DOCUMENT RESUME

ED 294 880 TM 011 241

AUTHOR Lad, F. R.; And Others
TITLE The Fundamental Theorem of Prevision. Technical Report No. 506. November 1987.
INSTITUTION Minnesota Univ., Minneapolis. School of Statistics.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE 5 Nov 87
NOTE 45p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Linear Programing; Mathematical Models; *Probability; *Statistical Analysis; Theories
IDENTIFIERS *Prevision

ABSTRACT

B. De Finetti's "Fundamental Theorem of Probability" is reformulated as a computable linear programming problem. The theorem is substantially extended, and shown to have fundamental implications for the theory and practice of statistics. It supports an operational meaning for the partial assertion of prevision via asserted bounds. The theorem is expanded to apply to general quantities; to allow bounds and orderings on previsions as input to the programming problem; and to yield bounds, even on conditional previsions, as output. Consequences include the ultimate strengthening of any probability inequality based on linear constraints, such as the Bienayme-P. L. Chebyshev inequality and an inequality related to Kolmogorov's inequality, but based only on the judgment of a sequence of quantities as exchangeable. Included in the wide variety of potential applications are the safety assessment of complex engineering systems, the analysis of agricultural production statistics, and a synthesis of subjective judgments in macro-economic forecasting. Prevision is explicitly recognized as a completion of the notion of logical assertion, introduced by Frege. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED294880

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

The Fundamental Theorem of Prevision

by

F.R. Lad[1], J.M. Dickey[2], and M.A. Rahman[3]

Technical Report No. 506
November 1987

UNIVERSITY OF MINNESOTA

SCHOOL OF STATISTICS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. A. Dickey
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

The Fundamental Theorem of Prevision

by

F.R. Lad[1], J.M. Dickey[2], and M.A. Rahman[3]

Technical Report No. 506
November 1987

Acknowledgements

This paper summarizes joint research by the authors, supported in its initial phase by the State University of New York at Albany, Department of Mathematics and Statistics and Department of Economics. We have profited from comments on earlier versions of this work presented at the Universities of Trieste, Rome, and Hull, and at the Special Studies Research Section of the Federal Reserve Board of Governors in Washington, D.C. Work by Professor Dickey was supported by the U.S. National Science Foundation, Research Grant Number DMS-8614793.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

[1] University of Canterbury, Christchurch, NZ

[2] University of Minnesota, Minneapolis, MN

[3] Trenton State College, Trenton, NJ

5 Nov. 1987

The Fundamental Theorem of Prevision

By F. R. LAD[1], J. M. DICKEY[2], and M. A. RAHMAN[3]

[1]University of Canterbury, Christchurch, NZ
[2]University of Minnesota, Minneapolis, MN
[3]Trenton State College, Trenton, NJ

SUMMARY

De Finetti's "Fundamental Theorem of Probability" is reformulated as a computable linear programming problem. The theorem is substantially extended, and shown to have fundamental implications for the theory and practice of statistics. It supports an operational meaning for the partial assertion of prevision via asserted bounds. We extend the theorem to apply to general quantities, to allow bounds and orderings on previsions as input to the programming problem, and to yield bounds, even on conditional previsions, as output. Consequences include the ultimate strengthening of any probability inequality based on linear constraints, such as the Bienaymé-Chebyshev inequality and an inequality related to Kolmogorov's inequality, but based only on the judgement of a sequence of quantities as exchangeable. Included in the wide variety of potential applications are the safety assessment of complex engineering systems, the analysis of agricultural production statistics, and a synthesis of subjective judgments in macroeconomic forecasting. In our discussion, prevision is explicitly recognized as a completion of the notion of logical assertion, introduced by Frege.

Keywords: LOGICAL DEPENDENCE; SUBJECTIVE PROBABILITY; COHERENCE; BOUNDS ON PREVISION; ORDINAL PROBABILITY; LINEAR PROGRAMMING; BIENAYME-CHEBYSHEV INEQUALITY; KOLMOGOROV'S INEQUALITY; EXCHANGEABILITY; LAWS OF LARGE NUMBERS; LOGICAL ASSERTION.

1. Introduction

Without elaborating on the choice of name for his theorem, de Finetti (1970, 3.10.1) announced as the "Fundamental Theorem of Probability" a derivation of bounds on the numerical assessment of the prevision of an event, bounds that are required by and insure its coherence with coherent previsions already asserted for N other events. The logic behind the theorem had already been presented in his Paris lectures, "Foresight: Its Logical Laws, Its Subjective Sources" (1937, Ch. 1), and the importance of the result had been recognized in the analysis of finite additivity in his paper "On the Axiomatization of Probability" (1949, 5.9). In the latter paper, the result is expressed in terms that identify the limitations under which a coherent prevision function specified over a linear space of events can be extended to a coherent function over a larger linear space. The analysis there is presented at such a level of mathematical abstraction that it has drawn scant attention. The technical prelude to the Fundamental Theorem in de Finetti (1970) is prolonged over at least 70 pages of introductory concepts and examples. Particularly important is the discussion of logical dependence, logical independence, and logical semi-dependence among events.

If a poll were taken of members of statistics societies throughout the world, we doubt that even 1 percent would say they considered "the fundamental theorem of probability" to be the result so designated by de Finetti. Even among statisticians who would call themselves "Bayesian", we doubt that the figure would reach 5 percent. In small groups of statisticians to whom we have addressed the question of identifying the fundamental theorem of probability, responses have ranged from "the Law of Large Numbers", to "the Central Limit Theorem", to "the Law of the Iterated Logarithm", to "There is no fundamental theorem of probability." A bold Bayesian would sometimes suggest Bayes' Theorem, or even de Finetti's theorem on the representation of exchangeable distributions.

The present paper is meant to elucidate the Fundamental Theorem in a constructive computable form, to extend it in useful ways, and to reveal its fundamental character by showing its comprehensive applicability and the resolution it provides for substantive issues in probability and statistics. After preliminary definitions and concepts (Section 2), we characterize the theorem as a linear programming problem (Section 3), first suggested by Bruno and Gilio

(1980) and extended by Rahman (1987). The Fundamental Theorem in linear programming form provides a computational procedure whereby any knowledge you actively assert via your previsions for N specific quantities enters as input into the program in terms of linear restrictions. The maximum and minimum of an objective function, computed as output from the program, serve as bounds on the prevision you may assert for a further specific quantity if it is to cohere with the N previsions you have already asserted as input. These are the narrowest such bounds. They guarantee the coherence of the full set of N+1 asserted previsions if the first N are themselves coherent.

After a careful discussion, we interpret the Fundamental Theorem of Probability to support the process of *asserting bounds* on previsions as an operationally meaningful representation of uncertain knowledge. With this interpretation, the theorem provides a standpoint for evaluating the controversial discussions of interval probabilities that have continued throughout this century in works such as Keynes (1921), Borel (1924), Koopman (1940), Reichenbach (1949), Good (1950), Smith (1961), de Finetti and Savage (1962), Scott (1964), Fishburn (1965,1985), Dempster (1967), Suppes (1974,1981), Shafer (1976), and Leamer (1986). (The list is not exhaustive.) We expand the Fundamental Theorem to allow assertions of bounds on incompletely assessed previsions as the primary input specifications of uncertain knowledge. Even more generally, assertions of mere orderings of prevision and other linear inequalities are shown to be meaningful inputs, with numerical implications computable within the linear programming framework.

Finally, we extend the theorem beyond the domain of events to a fundamental theorem of prevision for general quantities (Section 4). Any prevision inequality holding under linear equality or inequality constraints receives its strongest possible statement as a consequence of our general result. One corollary strengthens and completes the Bienaymé-Chebyshev inequality in the context of uncertainty about bounded discrete measurements. Another gives an inequality related to Kolmogorov's inequality, but involving quantities judged as exchangeable. A final extension has implications for cohering assertions of conditional previsions. The extension to conditional prevision requires a nonlinear programming computation, for which we provide a simple algorithm. The output bounds on conditional previsions have direct applicability in operational-subjective statistical methods.

Our results are illustrated by small-scale computations (Section 5). From the immense scope of potential practical applications, we suggest examples in engineering, agronomy, and macroeconomic forecasting. Concluding comments (Section 6) dwell on the logical category of prevision as an assertion, in the sense introduced by Frege (1879). In this light, we recognize the Fundamental Theorem of Prevision as a generalization of the deductive closure result of Hilbert and Ackermann (1938, I.§9.).

2. Preliminaries

Most of this section is a concise summary of concepts that are developed by de Finetti with extensive examples in chapters 2 and 3 of his treatise (1970). Readers who are not familiar with the de Finetti approach are asked to pay special attention to the definitions. Familiar sounding terms are often defined with a different meaning and syntax than in the measure-theoretic characterization of probability. For example, an event in the usual formulation is a set; whereas in our terminology, an event is a quantity, a number.

A *quantity*, X, is the numerical outcome of a particular operationally defined measurement. Hence, X is a well defined number, although its numerical value may be unknown at the time X is contemplated. The set of all numbers that are possible results of performing the operation is called the *realm* of the quantity, denoted by $\Re(X)$. Typically, it has a finite number of elements, called the *size* of the realm. The analysis in this paper is confined to the realistic case of a realm with finite size. A quantity, E, whose realm is $\Re(E) = \{0,1\}$ is called an *event*. If E is an event, then $\tilde{E} \equiv (1-E)$ is also an event. Definitional restrictions on events specify *logical relations* among them. For example, N events are said to be *incompatible* if their definitions imply that their sum cannot exceed 1. Similarly, N events are *exhaustive* if their sum cannot be less than 1. N events are said to *constitute a partition* if they are both incompatible and exhaustive, that is, if their sum necessarily equals 1. The individual events in this case are called *constituents* of the partition.

Any N events ($N \geq 1$) *generate* a partition with S(N) constituents. S(N) is called the *size* of the partition generated by $E_1, \ldots, E_N$. The constituents of this partition are those S(N) summands in the multiplicative expansion of the

expression $1 = \Pi_{i=1}^{N} (E_i + \tilde{E}_i)$ that are events. This is to say, their realms contain both (and only) the numbers 0 and 1. A typical summand in this expansion is a product of N events, such as $E_1\tilde{E}_2\tilde{E}_3E_4\cdots\tilde{E}_{N-1}E_N$. There are, of course, 2^N summands of this form. But some of them may not be events, since some of the summands necessarily equal zero if there are logical restrictions among the multiplicand events that generate the partition. For a simple example, suppose N=2, and $E_2 \equiv 1 - E_1$. Then neither E_1E_2 nor $\tilde{E}_1\tilde{E}_2$ are events, since they both necessarily equal 0. But both $E_1\tilde{E}_2$ and $\tilde{E}_1E_2$ are events. Thus, $S(2) = 2$, rather than 4. If every summand in the product expansion $\Pi_{i=1}^{N} (E_i + \tilde{E}_i)$ is an event, then $S(N) = 2^N$. Otherwise $S(N) < 2^N$. Throughout this paper, we will denote the constituents of the partition generated by the events $E_1, \ldots, E_N$ using the symbols $C_1, \ldots, C_{S(N)}$.

Geometrically, the S(N) constituents of the partition generated by N events can be represented by points in N-space, specifically, by S(N) designated vertices among the 2^N vertices of the N-dimensional unit cube. If there are no logical restrictions among the generating events, then $S(N)=2^N$, and every vertex of the N-dimensional cube represents a constituent of the partition. In this case we say the events are *completely logically independent*. But if there are any logical restrictions among $E_1, \ldots, E_N$, then some of the vertices must be removed from the N-dimensional cube in order to represent only the constituents of the partition generated by the N events. In such a case we say that the operational definitions of the events entail some degree of *logical dependence*. Figures 2.1 and 2.2 exhibit two possible configurations of logical dependence among three events. In Figure 2.1, the two events E_1 and E_2 are completely logically independent, while E_3 is their logical conjunction. It is defined functionally as the product $E_3 \equiv E_1E_2$. In Figure 2.2, the three events F_1, F_2, and F_3 are incompatible. Yet none of them is defined functionally in terms of the other two. De Finetti referred to such events as *logically semidependent*.

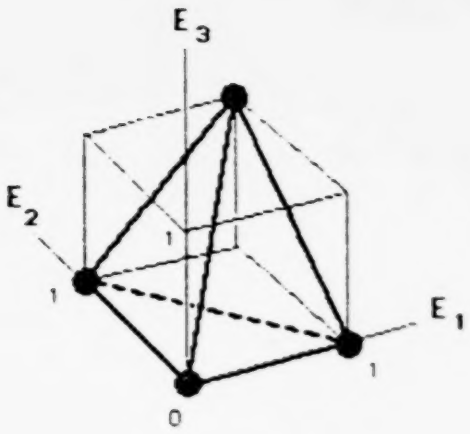

Figure 2.1. Logically Dependent Events. The two events E_1 and E_2 are completely logically independent, whereas event E_3 is their logical conjunction: $E_3 \equiv E_1E_2$.

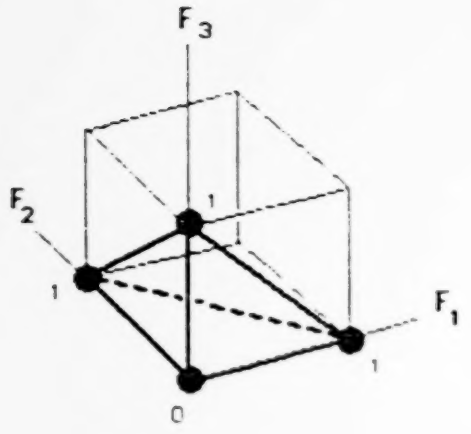

Figure 2.2. Logically Semidependent Events. The three events F_1, F_2, and F_3 are incompatible. Nevertheless, none of them is a logical function of the other two.

These concepts can be generalized to vectors. A vector of quantities, $\mathbf{X}_N \equiv (X_1, \ldots, X_N)^T$, is a vector whose components are quantities. The realm of such a vector, denoted by $\mathcal{R}(\mathbf{X}_N) \subset R^N$, is the set of vectors that represent possible outcome values obtained by performing the operations defining all the component quantities. The component quantities of $\mathbf{X}_N$ are said to be completely logically independent if $\mathcal{R}(\mathbf{X}_N)$ equals the cartesian product of the realms of its components. Otherwise the quantities are said to entail some degree of logical dependence. A vector of quantities generates a partition whose constituents are the events of the form $(\mathbf{X}_N = \mathbf{x}_N)$ where $\mathbf{x}_N$ is in $\mathcal{R}(\mathbf{X}_N)$. Thus, the size S(N) of the partition generated by N quantities equals the size of the realm of their vector. [Parentheses around a mathematical relation, such as $(\mathbf{X}_N = \mathbf{x}_N)$, should be taken to define an event equal to 1 if the relation holds, and equal to 0 otherwise.]

Your *prevision* for a vector of quantities $\mathbf{X} = (X_1, \ldots, X_N)^T$ is the vector of numbers $P(\mathbf{X}) \equiv (P(X_1), \ldots, P(X_N))^T$ you specify, with the understanding that you are thereby *asserting* your indifference to engaging any transaction that would yield you the net (sum of products) $\mathbf{s}^T[\mathbf{X} - P(\mathbf{X})]$ pounds sterling, where $\mathbf{s} = (s_1, \ldots, s_N)^T$ is any vector of scale constants. Your indifference must apply to

vectors $\mathbf{s}$ in every direction. It may be qualified only that the components of $\mathbf{s}$ must be sufficiently small that the net yield of any relevant transaction does not transgress the limited region over which your utilities are approximately linear. For example, you may stipulate that your assertion of indifference pertains only if $\mathbf{s}$ is scaled so that the maximum gain or loss you can incur from the yield $\mathbf{s}^T[\mathbf{X} - P(\mathbf{X})]$ is no greater than 10 pounds. (For detailed discussion of this feature, see de Finetti, 1970, 3.2.) If any component of $\mathbf{X}$ is an event, then the corresponding component of your prevision vector is called your *probability* for that event.

In asserting your prevision $P(\mathbf{X})$, you are avowing your willingness to buy *and* your willingness to sell a claim to $\mathbf{s}^T\mathbf{X}$ pounds in exchange for payment of $\mathbf{s}^T P(\mathbf{X})$ pounds. This is an operational implication of the stipulation that the vector $\mathbf{s}$ in the yield expression $\mathbf{s}^T[\mathbf{X} - P(\mathbf{X})]$ may have any direction. Having asserted your own $P(\mathbf{X})$, then for any vector $\mathbf{p}_1 \leq P(\mathbf{X})$, you would presumably also be willing to pay $\mathbf{s}^T\mathbf{p}_1$ pounds for a claim to $\mathbf{s}^T\mathbf{X}$ pounds where every component of $\mathbf{s}$ is positive. For this transaction would yield you at least as much as paying $\mathbf{s}^T P(\mathbf{X})$ pounds for a claim to $\mathbf{s}^T\mathbf{X}$ pounds. Similarly, for any vector $\mathbf{p}_2 \geq P(\mathbf{X})$, you would presumably be willing to sell a claim to $\mathbf{s}^T\mathbf{X}$ pounds in return for payment of $\mathbf{s}^T\mathbf{p}_2$ pounds.

Let us tarry a moment to highlight the technical aspect of defining prevision as an assertion you make regarding the value of $\mathbf{X}$. The realm of $\mathbf{X}$ presumably delineates all the various values of its component measurements that anyone can validly contemplate as possible: whereas your prevision $P(\mathbf{X})$ represents your operationally defined judgment of the value of $\mathbf{X}$ on the basis of such contemplation. (Someone else may assert a different value as his/her $P(\mathbf{X})$. Neither of you are estimating a "true" or "correct" value of $\mathbf{X}$, but rather asserting your own valuation of $\mathbf{X}$.) This distinction between $\mathcal{R}(\mathbf{X})$ and $P(\mathbf{X})$ parallels that introduced by Frege (1879) in mathematical logic. Within the confines of two-valued logic, he introduced notation to distinguish the content of a declarative sentence, which may be true or false, from a proposition, which is an assertion by someone that the sentence is true. The rules of two-valued logic govern the self-consistency of several propositions, requiring that you do not assert both the truth of a sentence, A, and the truth of its negation, $\tilde{A}$. The extension of these rules to the logic of uncertainty is motivated by the desirable property that your assertions of prevision be coherent.

Your prevision for a vector of quantities, $P(\mathbf{X})$, is said to be *coherent* as long as you do not assert by it your indifference to some transaction that would surely yield you a loss, no matter what the outcome value of $\mathbf{X}$ may be among the possibilities in $\mathfrak{R}(\mathbf{X})$. Algebraically, the coherency of your specified $P(\mathbf{X})$ requires that there exists no vector $\mathbf{s}$ with sufficiently small components for which, for some $\epsilon > 0$, $\mathbf{s}^T[\mathbf{x} - P(\mathbf{X})] < -\epsilon < 0$ for every vector $\mathbf{x} \in \mathfrak{R}(\mathbf{X})$. This specification of this requirement leads to the algebraic characterization of coherent prevision as a linear functional over the space of linear functions of $\mathbf{X}$. By a standard supporting-hyperplane argument, the set of all coherent vector previsions assessable for the vector of quantities $\mathbf{X}$ is identical to the convex hull of $\mathfrak{R}(\mathbf{X})$ in N-dimensional space. The coherent extendibility of your asserted linear functional, P, to larger spaces is the subject of the fundamental theorem of prevision, to be discussed.

Your *conditional prevision* for a quantity X conditional on E, denoted $P(X|E)$, is defined as the number you specify with the understanding that you are thereby asserting your indifference to engaging any transaction that would yield you the net gain of $s[XE - P(X|E)E]$ pounds sterling. Such a transaction is called a *contingent transaction* for X, contingent on E. For the yield from the transaction (gain or loss) will differ from 0 only if the event E in fact equals 1. A conditional prevision assertion $P(X|E)$ coheres with assertions of $P(XE)$ and $P(E)$ if and only if $P(XE) = P(X|E)P(E)$. This definition of conditional prevision makes no reference to any assertion of prevision you might make in the future. Your conditional prevision represents an operationally defined judgment you make now about the value of X and E, based on your current state of uncertain knowledge. (See Goldstein 1985 for discussion and developments based on this distinction.)

We conclude these preliminaries with the observation that any vector of events, $\mathbf{E}_N$, can be written as a linear function of the vector of constituents of the partition the events generate, $\mathbf{C}_{S(N)}$, via the equation

$$\mathbf{E}_N = \mathbf{R}_{N,S(N)}\,\mathbf{C}_{S(N)} \quad .$$

Here $\mathbf{R}_{N,S(N)}$ is the $[N \times S(N)]$ matrix whose columns are the vector elements of the realm $\mathfrak{R}(\mathbf{E}_N)$. Since every entry of $\mathbf{R}_{N,S(N)}$ equals either 0 or 1, each column vector of $\mathbf{R}_{N,S(N)}$ associates a specific constituent of the partition with some vertex of the N-dimensional unit cube. The equality of $\mathbf{E}$ and $\mathbf{RC}$ merely states

the identity of each event E_i with the sum of specific identifiable constituents of the partition generated by $E_1, \ldots, E_N$. These constituents are identified by expanding the right side of the equation, $E_i = E_i[\Pi_{j=1}^{N}(E_j + \breve{E}_j)(j \neq i)]$, and then recognizing the proscribed summands in the resulting expression that necessarily equal 0 due to logical restrictions among the events generating the partition. For example, the vector of three events whose realm is displayed in Figure 2.1 can be expressed as $\mathbf{E}_3 = \mathbf{R}_{3,4}\mathbf{C}_4$:

$$\begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} ,$$

where C_j is the event that the vector $\mathbf{E}_3$ equals column j of the matrix $\mathbf{R}_{3,4}$. Notice that the columns of $\mathbf{R}_{3,4}$ are also the vector elements of the realm $\mathcal{R}(\mathbf{E}_3)$, represented by bold dots in Figure 2.1. More generally, a similar equation characterizes any vector of quantities, $\mathbf{X}_N$, as

$$\mathbf{X}_N = \mathbf{R}(\mathbf{X}_N)\mathbf{C}_{S(N)} ,$$

where $\mathbf{R}(\mathbf{X}_N)$ is the matrix whose columns are the elements of the realm $\mathcal{R}(\mathbf{X}_N)$, and $\mathbf{C}_{S(N)}$ is the vector of constituent events $(\mathbf{X}_N = \mathbf{x}_N)$, one for each possible observation vector $\mathbf{x}_N$ in the realm $\mathcal{R}(\mathbf{X}_N)$.

Finally, notice that although the numerical values of the quantities $X_1, \ldots, X_N$ and of the constituents $C_1, \ldots, C_{S(N)}$ may well be unknown to you, you can be certain that the sum of the constituents equals 1. That is, $\Sigma\mathbf{C}_{S(N)} = 1$, since $C_1, \ldots, C_{S(N)}$ constitute a partition, by construction. [We use the notation $\Sigma\mathbf{v}$ for the sum of the components of a vector $\mathbf{v}$. We will also have recourse to denote by $\mathbf{1}_N$ the N-dimensional column vector with every component equal to 1.]

3. The Fundamental Theorem of Probability: Initial Extensions

The operational-subjective theory of probability allows you to assert, as your prevision (probabilities) for a vector of events, any vector of numbers you please, subject only to the restriction that your assertion be coherent. The coherency restriction will then *define* your prevision operator as a linear functional on the space of linear functions of the event vector. Notice that your prevision operator is *not* defined for all functions on the basis of some underlying measure. Rather, your prevision for a vector of quantities becomes defined only when you actively assert your willingness to engage the transactions specified in the definition. Coherency requires that when you assert this willingness, you concomitantly assert your willingness to engage in specified transactions involving linear combinations of the quantities, whose net yields would be identical to the yields from transactions you have expressly asserted to be acceptable. Now suppose you coherently specify your probabilities for a vector of N events, $\mathbf{E}_N$. De Finetti's fundamental theorem of probability characterizes the numerical restrictions on your assessment of prevision for *any* further event, E_{N+1}, that are required by -- and insure -- the coherency of your overall prevision for the vector of events $\mathbf{E}_{N+1} = (E_1, \ldots, E_N, E_{N+1})^T$. The first theorem we present is a reformulation of the fundamental theorem as a linear programming problem. It appears first to have been suggested in such a form by Bruno and Gilio (1980), while the subsequent extensions in this section were developed and discussed in the thesis of Rahman (1987).

Fundamental Theorem of Probability. Let $\mathbf{E}_N$ be a vector of events for which you have specified your prevision vector, $P(\mathbf{E}_N) = \mathbf{p}_N$; and let E_{N+1} be *any* other event. Depending on the logical relations among the events $E_1, \ldots, E_N, E_{N+1}$, they generate a partition of size $S(N+1) \leq 2^{N+1}$. Denote by $\mathbf{C}_{S(N+1)}$ the vector that comprises the constitutents of this partition. By construction, the vector $\mathbf{E}_{N+1} = \mathbf{R}_{N+1,S(N+1)}\mathbf{C}_{S(N+1)}$, for the appropriate matrix $\mathbf{R}_{N+1,S(N+1)}$. Denote the first N rows of $\mathbf{R}_{N+1,S(N+1)}$ by $\mathbf{R}_{N,S(N+1)}$, and the $(N+1)^{st}$ row by $\mathbf{r}_{N+1}$. Then, for the coherency of an extended prevision assertion for all components of $\mathbf{E}_{N+1}$, $P(\mathbf{E}_{N+1}) = (\mathbf{p}_N^T, P(E_{N+1}))^T$, it is both necessary and sufficient that the numerical value of your $P(E_{N+1})$ lie within the interval $[l_{N+1}, u_{N+1}]$, where the values of l_{N+1}

and u_{N+1} are determined by solving the following two linear programming problems:

Find those $S(N+1)$-tuples $\mathbf{q}_{S(N+1)} = (q_1, q_2, \ldots, q_{S(N+1)})^T$ that yield the extrema,

$$l_{N+1} \equiv \text{minimum } (\mathbf{r}_{N+1}\mathbf{q}_{S(N+1)}) \quad \text{and} \quad u_{N+1} \equiv \text{maximum } (\mathbf{r}_{N+1}\mathbf{q}_{S(N+1)}),$$

both subject to the $(N+1)$ linear equality constraints

$$\mathbf{R}_{N,S(N+1)}\,\mathbf{q}_{S(N+1)} = \mathbf{p}_N \quad \text{and} \quad \Sigma\,\mathbf{q}_{S(N+1)} = 1,$$

along with the $S(N+1)$ non-negativity restrictions that each component of $\mathbf{q}_{S(N+1)}$ be non-negative. The feasible region for these programming problems is empty if and only if your original assertion of $P(\mathbf{E}_N) = \mathbf{p}_N$ is incoherent.

Proof. An assertion $P(\mathbf{E}_{N+1}) = \mathbf{p}_{N+1}$ is coherent if and only if the vector $\mathbf{p}_{N+1}$ lies within the convex hull of the set $\mathcal{R}(\mathbf{E}_{N+1})$. Now the event vector $\mathbf{E}_{N+1}$ is a linear transformation of the constituent vector $\mathbf{C}_{S(N+1)}$ it generates. The transformation takes vectors in $S(N+1)$-dimensional space into $(N+1)$-dimensional space by the transforming matrix $\mathbf{R}_{N+1,S(N+1)}$, viz., $\mathbf{E}_{N+1} = \mathbf{R}_{N+1,S(N+1)}\mathbf{C}_{S(N+1)}$. Under this transformation, the convex hull of $\mathcal{R}(\mathbf{E}_{N+1})$ is the image of $\mathcal{R}(\mathbf{C}_{S(N+1)})$. Thus, the vector $\mathbf{p}_{N+1}$ lies within the convex hull of $\mathcal{R}(\mathbf{E}_{N+1})$ if and only if it can be obtained by the same linear tranformation of some vector within the convex hull of the realm $\mathcal{R}(\mathbf{C}_{S(N+1)})$. Since the components of $\mathbf{C}_{S(N+1)}$ constitute a partition, the convex hull of $\mathcal{R}(\mathbf{C}_{S(N+1)})$ is the simplex of vectors $\mathbf{q}_{S(N+1)} = (q_1, \ldots, q_{S(N+1)})^T$ whose components are nonnegative and sum to 1. The assertion $P(\mathbf{E}_{N+1}) = \mathbf{p}_{N+1}$ is an extension of the assertion $P(\mathbf{E}_N) = \mathbf{p}_N$ if and only if the first N components of the vector $\mathbf{p}_{N+1}$ are identical to the components of $\mathbf{p}_N = \mathbf{R}_{N,S(N+1)}\mathbf{q}_{S(N+1)}$ for some qualifying vector $\mathbf{q}_{S(N+1)}$. Thus, satisfaction of the linear programming formulation is both necessary and sufficient for an assertion $P(\mathbf{E}_{N+1}) = \mathbf{p}_{N+1}$ to be a coherent extension of the assertion $P(\mathbf{E}_N) = \mathbf{p}_N$. The same logic underlies the final statement in the theorem, that the original assertion $P(\mathbf{E}_N) = \mathbf{p}_N$ is incoherent if and only if the feasible region of the specified programming problems is empty. ∇

Let us make a few simple observations before a deeper discussion.

At one extreme, if E_{N+1} happens to be a linear function of $E_1,\ldots,E_N$, then $P(E_{N+1})$ is determined exactly, on account of the linearity property of coherent prevision. In this case $l_{N+1} = u_{N+1} = P(E_{N+1})$. At the other extreme, if E_{N+1} happens to be completely logically independent of $E_1,\ldots,E_N$, that is, if $S(N+1) = 2S(N)$, that is, if E_{N+1} and $\tilde{E}_{N+1}$ are both compatible with every constituent of the partition generated by $E_1,\ldots,E_N$, then $l_{N+1} = 0$ and $u_{N+1} = 1$. In this case, the boundaries on the coherent assertion of $P(E_{N+1})$, as an extension of the assertion $P(\mathbf{E}_N) = \mathbf{p}_N$, are not affected at all by the specific components of the vector $\mathbf{p}_N$. (A coherent prevision assessment for any event, of course, must lie within the interval [0,1].)

Between these two extremes lie all the intermediate possibilities of logical dependence conceivable among $E_1,\ldots,E_{N+1}$. The tightness of the bound on $P(E_{N+1})$ depends on the numerical values of $P(E_1),\ldots,$ and $P(E_N)$ as well as on the logical relations among $E_1,\ldots,E_{N+1}$. For example, notice that in Figure 2.1 if $P(E_1) = P(E_2) = .5$, then the bounds on $P(E_3)$ are 0 and .5. For any value of $P(E_3)$ outside these bounds, the vector $P(\mathbf{E}_3) = (.5,.5,P(E_3))$ would lie outside the convex hull of the realm $\mathfrak{R}(\mathbf{E}_3)$, outlined in bold. Whereas, if $P(E_1) = P(E_2) = .7$, then the bounds on $P(E_3)$ are .4 and .7. Within the convex hull of the four possible outcome vectors, the convex hull of $\mathfrak{R}(\mathbf{E}_3)$, all vectors that project orthogonally onto the point $(p_1,p_2) = [P(E_1),P(E_2)]$ lie within the bounds specified by the two linear programming problems.

The major practical difference between de Finetti's characterization of coherent prevision as a linear functional and the more common measure-theoretic axiomatization of probability can be seen by comparing this fundamental theorem with a corresponding axiom of the usual approach. The measure-theoretic conception supposes that a unique probability measure is defined on every "elementary event", that is, a set corresponding to a constituent of our partition generated from $E_1,\ldots,E_{N+1}$. Then it is axiomatic that the probability of any union of these disjoint events [note the measure-theoretic and set-theoretic language] equals the sum of the probabilities of the elementary events in the union. Bayesian statistical theorists who have attempted to use this mathematical

formulation with a subjective interpretation are justly criticized by the objecting practitioner who questions "How can I possibly assess my probability for each of those elementary events?" For S(N+1) can be much larger than N+1, even as large as 2^{N+1}. The characterization of coherent prevision as a linear functional allows you, as the practitioner, to assess your prevision for as many or as few events as you feel able and interested. Notice that any vector $\mathbf{q}_{S(N+1)}$ satisfying the linear programming constraints would be coherent, and would cohere with the assertion $P(\mathbf{E}_N) = \mathbf{p}_N$, if it were asserted as a prevision of the constitutent vector $\mathbf{C}_{S(N+1)}$. The usefulness of the fundamental theorem of probability lies in the fact that the logical relations among the events of interest to you can be exploited in aiding your assessment of $P(E_{N+1})$, without the necessity that you identify your prevision for every constituent of the partition generated by $E_1, \ldots, E_{N+1}$.

3.1. *Discussion: Bounds on Prevision at the Base of the Assessment Process*

After you have coherently asserted your prevision $P(\mathbf{E}_N) = \mathbf{p}_N$, the requirement of the fundamental theorem that $l_{N+1} \leq P(E_{N+1}) \leq u_{N+1}$ has two practical implications. One is cautionary. The other is behavioural. As a guideline, the requirement cautions that if you now undertake to specify your $P(E_{N+1})$, it had better lie within the interval $[l_{N+1}, u_{N+1}]$, or else you will have expressed an incoherent opinion. If you desire to be coherent, a reassessment of $P(E_1), \ldots, P(E_N)$ would be in order if you are satisfied with your assertion of $P(E_{N+1})$ outside of the interval $[l_{N+1}, u_{N+1}]$. Indeed, this is the language in which the fundamental theorem has been stated. But in addition, the theorem already has a *behavioural consequence* for you, even if you never assert a prevision value for E_{N+1}. The theorem implies that the coherency of your prevision operator along with the logical relation of E_{N+1} to $\mathbf{E}_N$ and your already specified assertion of $P(\mathbf{E}_N)$, together, amount to your avowed willingness to pay any amount up to (sl_{N+1}) for a claim to the unknown value (sE_{N+1}). [As noted in the preliminaries, s is qualified to be a small or moderate amount, say 10 pounds sterling.] For a combination of transactions involving only components of $\mathbf{E}_N$ can be arranged that will *surely not* return you *more* than (sE_{N+1}) and for which you have already asserted your willingness to pay (sl_{N+1}). Similarly, you are avowedly willing to offer for sale a claim to (sE_{N+1}) in return for at least (su_{N+1}).

This statement of behavioural implications for de Finetti's Fundamental Theorem of Probability is related operationally to Leamer's (1986) suggestion that a "bid-ask spread" be considered the basic meaningful unit for expressing one's uncertainty about a quantity within the operational-subjective framework. Although we do not subscribe to the entire argument presented in Leamer's paper, his operational meaning for asserting a probability interval is compelling. A much discussed criticism of the operational-subjective theory of probability hinges on the requirement that you specify a single price at which you are both willing to "buy" and willing to "sell" a quantity, in order that the theory have any content. The behavioural interpretation of the fundamental theorem softens this requirement. It is operationally meaningful to make a *partial assertion* of your prevision for a quantity X -- that your P(X) lies within the interval $[p_l, p_u]$. Formally, you thereby avow your willingness to engage any transaction that would yield you the net gain of $s_l[X - p_l] + s_u[p_u - X]$, so long as s_l and s_u are non-negative scalars small enough that your net gain or loss cannot be too large. Requiring coherency of a partial assertion of prevision, that you neither assert a willingness to accept a sure loss, nor a willingness to forego a sure gain, implies minimally that a coherent prevision interval $[p_l, p_u]$ must satisfy the inequalities: $\min \mathcal{R}(X) \leq p_l \leq p_u \leq \max \mathcal{R}(X)$.

In higher dimensions, this characterization of a partial assertion as the assertion of a prevision interval expands not merely to a prevision hyperinterval, but to a prevision polytope, perhaps highly irregular in shape. This follows from the fact that when you assert your willingness to engage in several individual transactions, coherence requires your willingness also to engage them in linear combinations (subject to the qualification that the scale of the net gain or loss not be too large). Moreover, a partial assertion regarding an individual quantity may be redundant in the context of other partial assertions you make. These ideas are presented most simply by an example.

Suppose that E_1 and E_2 are incompatible events, and that event E_3 is defined as their sum: $E_3 \equiv E_1 + E_2$. Thus, the convex hull of $\mathcal{R}(\mathbf{E}_3)$ is the plane triangle connecting the points (0,0,0), (1,0,1), and (0,1,1). This hull is depicted in Figure 3.1, projected onto the 2-dimensional space containing $\mathcal{R}(\mathbf{E}_2)$. Now, suppose further that you make the three partial assertions of prevision, $P(E_1) \in [.25, .5]$, $P(E_2) \in [.2, .3]$, and $P(E_3) \in [.5, .9]$. The dark polygon within the convex hull contains all the vectors in 2-dimensional space that satisfy the

restrictions specified by your several partial assertions. For any price vector (p_1, p_2) outside this polygon, you have effectively asserted your willingness to engage some transactions that involve buying or selling E_1 for p_1 *and/or* buying or selling E_2 for p_2. But you have not yet made any assertion of your position on exchanges involving prices represented by any vector within the polygon.

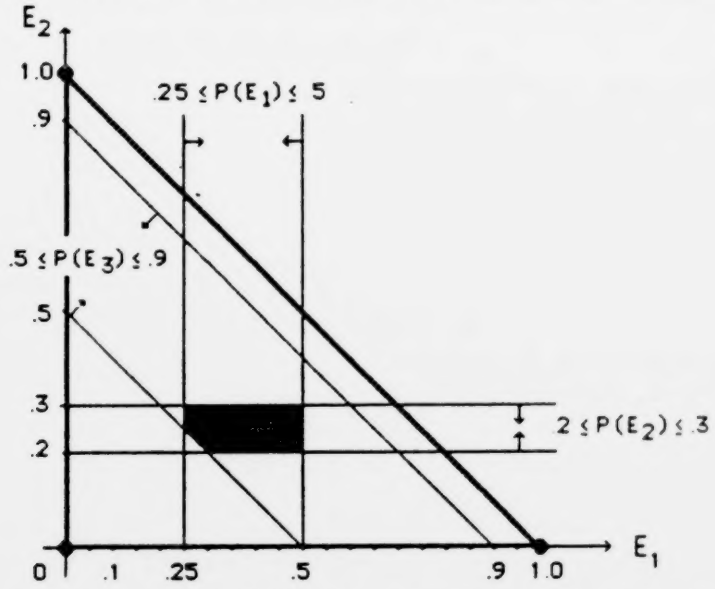

Figure 3.1. A partially asserted prevision polytope. The events E_1 and E_2 are incompatible, and $E_3 \equiv E_1 + E_2$. The convex hull of $\mathcal{R}(E_3)$, projected onto the 2-dimensional space containing $\mathcal{R}(E_2)$, is the heavily outlined half unit-square. The dark polygon within this convex hull is the partially asserted prevision polytope specified by the three partial assertions of prevision, $P(E_1) \in [.25, .5]$, $P(E_2) \in [.2, .3]$, and $P(E_3) \in [.5, .9]$.

There are two special features to note in this example. First is that the asserted upper bound, $P(E_3) \leq .9$, is redundant in light of the other two assertions of $P(E_1) \leq .5$ and $P(E_2) \leq .3$. For the willingness the latter signify, to engage in any transaction yielding $s_{1U}(.5 - E_1) + s_{2U}(.3 - E_2)$ as long as s_{1U} and s_{2U} are non-negative, implies a willingness to engage in any transaction yielding $s_{3U}(.8 - E_1 - E_2) = s_{3U}(.8 - E_3)$, signified by the assertion $P(E_3) \leq .8$. The second feature to note is that the assertion $P(E_3) \geq .5$ signifies a willingness to engage a transaction yielding $s_{3I}(E_3 - .5) = s_{3I}(E_1 + E_2 - .5) = s_{3I}(E_1 - .27) + s_{3I}(E_2 - .23)$, for example, even though .27 exceeds the lower partial assertion value of $P(E_1)$, and .23 exceeds the lower partial assertion value of $P(E_2)$. Thus, the vector (.27, .23) lies outside the polygon of partially asserted prevision.

The fundamental theorem of probability supports and even motivates the point of view that your intervals of *partial assertion* of prevision for individual quantities has a definite operational meaning in the representation of uncertain knowledge. The coherency requirement that you neither willingly accept sure losses nor willingly forego sure gains characterizes a *partially asserted prevision polytope*, the set of vectors that satisfy the inequalities of *all* your partial assertions, as a convex polytope lying within the convex hull of the realm of the quantity vector. You can be said to have asserted your prevision for a quantity, as defined in the preliminaries of this paper, only in the extreme case that your asserted prevision interval for that quantity consists of a single number. The fundamental theorem actually requires proponents of the operational-subjective formulation of uncertain knowledge to admit this viewpoint. For whatever precise prevision assertions you make for whatever quantities, the theorem shows us how to identify another quantity for which your avowed assertions are equivalent to a partial assertion.

The terminology *partial assertion of prevision* for the statement $P(X) \in [p_l, p_u]$ is expressly meant to connote that, conceivably, you can complete an assertion of your prevision for this quantity by a process of further introspection and sharper decision. Would you rather own a claim to X pounds or a claim to $(p_l + p_u)/2$ pounds? Once you decide, you will have strengthened your partial assertion of prevision either to $P(X) \in [(p_l + p_u)/2, p_u]$ or to $P(X) \in [p_l, (p_l + p_u)/2]$, depending on the decision. However, there are many useful ways you might decide to spend your time. So there can be no *requirement* that you assert a resolution of any particular value question such as this one. Several contemporary proponents of "interval probabilities" argue that probabilities are best considered to be *irreducible* intervals. Subjectivist proponents of this view say that "when I assert $P(E) \in [p_l, p_u]$, I mean that I would pay up to p_l for a claim to E, and I would sell a claim to E for p_u or more. But at prices between p_l and p_u, I will neither buy nor sell a claim to E." We do not subscribe to this viewpoint. Without further discussion here, let us merely state that such a position neglects the linearity of utility presumed in the qualification that the scale be small for the net yields from any relevant transactions.

3.2. *Extensions of the Theorem of Probability*

The fundamental theorem of probability can be extended to describe the implications of coherency for your partial assertion of probability intervals. The theorem, in the form stated above, makes only limited use of the rich possibilities of the linear programming structure. The constraint $\Sigma \mathbf{q}_{S(N+1)} = 1$, along with the $S(N+1)$ restrictions that each component of $\mathbf{q}_{S(N+1)}$ be non-negative, together specify the feasible region of vectors $\mathbf{q}_{S(N+1)}$ as the convex hull of the realm $\mathcal{R}(\mathbf{C}_{S(N+1)})$. The matrix $\mathbf{R}_{N+1,S(N+1)}$ transforms these vectors into $(N+1)$-dimensional space. Thus, in effect, these $[S(N+1)+1]$ restrictions on $\mathbf{q}_{S(N+1)}$ define a convex polytope in $(N+1)$-dimensional space. Each of the further N exact linear constraints specified by the equation $P(\mathbf{E}_N) = \mathbf{p}_N = \mathbf{R}_{N,S(N+1)}\mathbf{q}_{S(N+1)}$ reduces by 1 the dimension of the transformed feasible region. When all constraints are met, the coherent assertions $P(\mathbf{E}_{N+1})$ that are extensions of the assertion $P(\mathbf{E}_N) = \mathbf{p}_N$ are restricted to lie along a bounded one-dimensional line segment. Its endpoints are defined by the extrema of the designated linear programming problems.

We say this is merely "limited use" of the linear programming setup, since you need not go so far as to assert fully your prevision vector $P(\mathbf{E}_N)$ in order to compute numerical bounds for $P(\mathbf{E}_{N+1})$ with a linear programming algorithm. A computable solution of bounds for coherent assertions regarding $\mathbf{E}_{N+1}$ can still be achieved on the basis of partial assertions, $\mathbf{l}_N \leq P(\mathbf{E}_N) \leq \mathbf{u}_N$. Although these assertions may not reduce the *dimension* of your prevision polytope for $\mathbf{E}_{N+1}$, they could reduce its volume considerably. This is the tack we follow in stating our first extension of the fundamental theorem. (Its proof is contained informally in the preceding discussion.)

The Fundamental Theorem of Probability - Extension 1. Let $\mathbf{E}_N$ be any vector of events for which you make the partial assertions $\mathbf{l}_N \leq P(\mathbf{E}_N) \leq \mathbf{u}_N$. And let $\mathbf{E}_{N+1}$ be any other event. The logical relations among components of $\mathbf{E}_{N+1}$ specify that $\mathbf{E}_{N+1} = \mathbf{R}_{N+1,S(N+1)}\mathbf{C}_{S(N+1)}$. (Again, let $\mathbf{R}_{N,S(N+1)}$ denote the matrix composed of the first N rows of $\mathbf{R}_{N+1,S(N+1)}$, and let $\mathbf{r}_i$ denote the i^{th} row.) The coherency of your explicit assertions regarding the vector $\mathbf{E}_{N+1}$ entails that you

also avow the partial assertion $l_{N+1} \leq P(E_{N+1}) \leq u_{N+1}$, where l_{N+1} and u_{N+1} are determined by the solutions to the following two linear programming problems:

Find those $S(N+1)$-tuples $\mathbf{q}_{S(N+1)} = (q_1, \ldots, q_{S(N+1)})^T$ that yield the extrema

$$l_{N+1} \equiv \text{minimum } (\mathbf{r}_{N+1}\mathbf{q}_{S(N+1)}) \quad \text{and} \quad u_{N+1} \equiv \text{maximum } (\mathbf{r}_{N+1}\mathbf{q}_{S(N+1)}) ,$$

both subject to the linear constraints that

$$\mathbf{R}_{N,S(N+1)}\,\mathbf{q}_{S(N+1)} \geq \mathbf{l}_N ,$$

$$\mathbf{R}_{N,S(N+1)}\,\mathbf{q}_{S(N+1)} \leq \mathbf{u}_N , \text{ and}$$

$$\Sigma\, \mathbf{q}_{S(N+1)} = 1,$$

along with the non-negativity restrictions on the components of $\mathbf{q}_{S(N+1)}$. Moreover, the coherency of your several assertions about $\mathbf{E}_{N+1}$ defines your prevision polytope for $\mathbf{E}_{N+1}$ as the feasible region in these linear programming problems, transformed into $(N+1)$-dimensional space by the matrix $\mathbf{R}_{N+1,S(N+1)}$. Thus, for each component event E_i of $\mathbf{E}_{N+1}$, you avow, in effect, the partial assertion $l_i^* \leq P(E_i) \leq u_i^*$, where l_i^* and u_i^* are the extreme values attainable by the function $\mathbf{r}_i\,\mathbf{q}_{S(N+1)}$ within the feasible region. Δ

This form of the theorem exhibits the interconnections among all your partial assertions of prevision that are required by coherency. Your prevision for *each* of the N+1 events is constrained in the same fashion, by a bounding interval. The vector of your previsions for *all* of the N+1 events must lie within a convex polytope, the transformed feasible region of the programming problems. Any further decisive introspection motivating you to narrow one of your asserted intervals, $[l_i^*, u_i^*]$, could have an effect on the implied bounds for any or all other quantities, narrowing the associated intervals. For your explicit narrowing of the interval $[l_i^*, u_i^*]$ (for example, asserting $P(E_i)$ precisely) would amount to a more restrictive specification of the feasible region of vectors $\mathbf{q}_{S(N+1)}$ that are allowed by the programming problems.

Note that the implied intervals $[l_i^*, u_i^*]$ are "marginal" rather than "joint" intervals, in the sense that, because they are merely one-dimensional projections of the partial prevision polytope, their cartesian-product hyperinterval need not consist of points that would be coherent if asserted as prevision vectors. They are necessary but not jointly sufficient as bounds for coherent prevision vectors. The smaller partial prevision polytope is the set of all the coherent candidate prevision vectors.

A second useful extension of the fundamental theorem is readily apparent. In the linear programming context, a mere assertion of orderings among your previsions for several quantities or for linear combinations of them is sufficient to generate computable bounds that express your uncertain knowledge regarding any quantity. For example, you might assert that your $P(E_1) \geq P(E_2)$, meaning that you avow a willingness to exchange a claim to sE_2 pounds in return for a claim to sE_1 pounds (presuming s is not large). With similar operational meaning, you might assert that your $P(E_3)+P(E_4) \geq P(E_5)$, or even that your $P(E_6)+2P(E_7) \geq P(E_8)$. Moreover, any assertion of conditional prevision can be expressed as a linear constraint as well. A coherent assertion that your $P(E_1|E_2) = p_{1\cdot 2}$, for example, is equivalent to the assertion that $P(E_1E_2) = p_{1\cdot 2}P(E_2)$, which is to say, $P(E_1E_2) - p_{1\cdot 2}P(E_2) = 0$. This is a linear restriction on your prevision for the events E_2 and $E_3 \equiv E_1E_2$. Similarly, the partial assertion $P(E_1|E_2) \in [a,b]$ is representable by linear restrictions: $aP(E_2) - P(E_1E_2) \leq 0$, and $bP(E_2) - P(E_1E_2) \geq 0$. Each such statement is readily translated into linear constraints allowable in the linear programming framework: $\mathbf{a}\mathbf{R}_{N,S(N+1)}\mathbf{q}_{S(N+1)} \leq b$, for a suitably defined row vector $\mathbf{a}$ and an appropriate number b. (Without loss of generality, we will henceforth express all inequality assertions in such a "less than or equal to" form.) Let us merely state this second extension of the fundamental theorem in a summary fashion.

Fundamental Theorem of Probability - Extension 2. The fundamental theorem of probability extends further to allow meaningful partial assertions of prevision in the form $\mathbf{A}_{K,N}P(\mathbf{E}_N) \leq \mathbf{b}_K$ as input to the linear programming problems, and to imply computable bounds on coherent prevision for any linear combination of constituents, $P(\mathbf{r}\,\mathbf{C}_{S(N+1)})$. Δ

This extension of the fundamental theorem unifies the numerical representation of subjective probability with ideas of merely ordinal probability, as advanced in several works of Shackle (1949, 1955). According to de Finetti (1965), ideas behind such an extension were already underlying works in educational testing by Coombs, Milholland, and Womer (1956), Willey (1960), Chernoff (1961, 1962), and Dell'Era (1963).

We can summarize the position to which the fundamental theorem of probability has led us. The requirement of coherency provides that whatever knowledge you assert about a vector of events, no matter how meagre or how

detailed the knowledge may be, delineates a convex polytope that represents your prevision to the extent to which you have specified it. We need not presume that the volume of the polytope is reduced to zero by any *precise* specification of your prevision. Yet there is positive operational meaning to the knowledge you do specify.

4. The Fundamental Theorem of Prevision

Since events are merely quantities whose realm is {0,1}, it should not be surprising that the fundamental theorem of probability, and each of the extensions we have presented above, depicts a special case of a theorem applicable to prevision for general quantities. What may be surprising is the breadth of important results in statistical theory that are particular instances of the general result. We will state and prove the fundamental theorem of prevision in two parts. The first part is a comprehensive generalization of results we have already discussed. The second part reveals the bounds implied for coherent conditional prevision. After an intermediate discussion, we will dwell on two important corollaries.

In what follows, we presume $\mathbf{X}_N = (X_1, \ldots, X_N)^T$ to be a quantity vector, with a finite discrete realm $\mathcal{R}(\mathbf{X}_N)$ having S(N) members. We noted at the end of our preliminaries that $\mathbf{X}_N$ can be represented in terms of the linear equation

$$\mathbf{X}_N = \mathbf{R}_{N,S(N)}\mathbf{C}_{S(N)} \quad ,$$

where $\mathbf{R}_{N,S(N)} = \mathbf{R}(\mathbf{X}_N)$ is the $(N \times S(N))$ matrix whose columns are the vector elements of the realm $\mathcal{R}(\mathbf{X}_N)$, and $\mathbf{C}_{S(N)}$ is the $(S(N) \times 1)$ vector of constituent events of the form $(\mathbf{X}_N = \mathbf{x}_N)$, one for each element vector $\mathbf{x}_N$ in the realm $\mathcal{R}(\mathbf{X}_N)$. Individual rows of $\mathbf{R}_{N,S(N)}$ are denoted $\mathbf{r}_1, \ldots, \mathbf{r}_N$. Using the generalization of prevision to a prevision interval (operationally defined by the assertion of a bid-ask spread) and the generalization to the assertion of any preference representable by $\mathbf{a}^T P(\mathbf{X}_N) \leq b$, we can represent any knowledge you would like to assert about components of $\mathbf{X}_N$ via linear relations of the form $\mathbf{A}_{K,N} P(\mathbf{X}_N) \leq \mathbf{b}_K$. Based upon the characterization of coherent assertions as the foregoing of any sure losses and the accepting of all sure gains, we can now state simply and generally:

The Fundamental Theorem of Prevision, Part I. Let $\mathbf{X}_N$ be any vector of quantities for which you have partially asserted your prevision via the specifications $\mathbf{A}_{K,N}P(\mathbf{X}_N) \leq \mathbf{b}_K$. (The number K may be less than, equal to, or greater than N.) Then coherency implies that for any component, X_i, you assert, in effect, $P(X_i) \in [l_i, u_i]$, where the numerical values of l_i and u_i are calculated as the extreme values of the objective functions in the linear programming problems:

Find the two S(N)-tuples $\mathbf{q}_{S(N)} = (q_1, \ldots, q_{S(N)})^T$ that characterize

$$l_i \equiv \text{minimum } \mathbf{r}_i \mathbf{q}_{S(N)} \quad \text{and} \quad u_i \equiv \text{maximum } \mathbf{r}_i \mathbf{q}_{S(N)} \ ,$$

both subject to the linear constraints

$$\mathbf{A}_{K,N} \mathbf{R}_{N,S(N)} \mathbf{q}_{S(N)} \leq \mathbf{b}_K \ , \quad \text{and}$$

$$\Sigma \mathbf{q}_{S(N+1)} = 1 \ ,$$

along with the non-negativity restrictions on the components of $\mathbf{q}_{S(N)}$. The common feasible region for these programming problems, translated into N-dimensional space via the matrix $\mathbf{R}_{N,S(N)}$, constitutes your coherent prevision polytope for $\mathbf{X}_N$. This feasible region is non-empty if and only if your original assertion $\mathbf{A}_{K,N}P(\mathbf{X}_N) \leq \mathbf{b}_K$ is coherent.

Proof of Part I. This part of the theorem follows immediately from the second extension of the fundamental theorem of probability discussed in the previous section. For any general quantity can be represented as a linear combination of events: $X = \Sigma\, x_i(X = x_i)$, where the summation extends over all the possible observations x_i in the realm $\mathfrak{R}(X)$. A linear programming algorithm will necessarily yield finite extreme value solutions to these problems as long as the feasible region is not empty, since the feasible region is bounded. ∇

It is worth mentioning explicitly the reminder that the assertion of each individual $P(X_i)$ within its associated interval $[l_i, u_i]$ is necessary but not sufficient for the coherency of a prevision vector $P(\mathbf{X}_N)$. The necessary and sufficient condition for the coherence of the prevision vector $P(\mathbf{X}_N)$ is that it lie within the feasible region for these programming problems, translated into N-dimensional space via the matrix $\mathbf{R}_{N,S(N)}$.

Remember that any assertion of conditional prevision, such as the partial assertion $P(X|E) \in [a,b]$, can be incorporated into the form of input to the programming problems specified in this theorem. Due to the coherency requirement that $P(XE) = P(X|E)P(E)$, it is equivalent to the two assertions, $aP(E) - P(XE) \leq 0$ and $bP(E) - P(XE) \geq 0$. However, bounds on coherent conditional previsions cannot be computed as output from the theorem as stated, since cohering $P(X|E)$ is not a *linear* function of $P(XE)$ and $P(E)$. Indeed, we know, at least when you assert $P(E) > 0$, that $P(X|E)$ must equal the quotient $P(XE)/P(E)$. We can use this fact to derive a sufficient condition for the coherence of a conditional prevision as an extension of assertions $\mathbf{A}_{K,N} P(\mathbf{X}_N) \leq \mathbf{b}_K$.

For clarity in stating Part II of the fundamental theorem of prevision, we will refer to a further assertion of conditional prevision beyond the assertions $\mathbf{A}_{K,N} P(\mathbf{X}_N) \leq \mathbf{b}_K$ as a statement involving $P(X_{N+1}|X_{N+2})$, where X_{N+1} is a quantity and X_{N+2} is an event, denoted distinctly from the components of $\mathbf{X}_N$. You should be aware, however, that there is nothing special about these quantities. They could well both be components of $\mathbf{X}_N$ about which you have explicitly made partial assertions of your prevision.

The Fundamental Theorem of Prevision, Part II. Let $\mathbf{X}_N$ be any vector of quantities for which you have partially asserted your prevision via the specification, $\mathbf{A}_{K,N} P(\mathbf{X}_N) \leq \mathbf{b}_K$, as in Part I. Now let X_{N+1} and X_{N+2} be any other quantity and any event, respectively, and let X_{N+3} be defined as their product, $X_{N+3} = X_{N+1}X_{N+2}$. Supposing $\mathcal{R}(\mathbf{X}_{N+3})$ has $S(N+3)$ members, $\mathbf{X}_{N+3}$ is representable via the equation $\mathbf{X}_{N+3} = \mathbf{R}_{N+3,S(N+3)}\mathbf{C}_{S(N+3)}$. Let $\mathbf{R}_{N,S(N+3)}$ denote the matrix composed of the first N rows of $\mathbf{R}_{N+3,S(N+3)}$, and denote the final three rows of $\mathbf{R}_{N+3,S(N+3)}$ by $\mathbf{r}_{N+1}$, $\mathbf{r}_{N+2}$, and $\mathbf{r}_{N+3}$, respectively. Any further assertion of conditional prevision $P(X_{N+1}|X_{N+2})$ coheres with $\mathbf{A}_{K,N} P(\mathbf{X}_N) \leq \mathbf{b}_K$ if it lies within the interval $[l_{N+1|N+2}, u_{N+1|N+2}]$, where the numerical values $l_{N+1|N+2}$ and $u_{N+1|N+2}$ are calculated as extreme values of the objective functions in the nonlinear programming problems:

Find those $S(N+3)$-tuples, $\mathbf{q}_{S(N+3)} = (q_1, \ldots, q_{S(N+3)})^T$, that characterize

$$l_{N+1|N+2} \equiv \text{minimum } [\mathbf{r}_{N+3}\mathbf{q}_{S(N+3)} / \mathbf{r}_{N+2}\mathbf{q}_{S(N+3)}] \quad , \text{ and}$$

$$u_{N+1|N+2} \equiv \text{maximum } [\mathbf{r}_{N+3}\mathbf{q}_{S(N+3)} / \mathbf{r}_{N+2}\mathbf{q}_{S(N+3)}] \quad ,$$

both subject to the linear constraints

$$\mathbf{A}_{K,N}\,\mathbf{R}_{N,S(N+3)}\,\mathbf{q}_{S(N+3)} \leq \mathbf{b}_K \, , \quad \text{and}$$

$$\Sigma \mathbf{q}_{S(N+3)} = 1 \quad ,$$

along with the non-negativity restrictions on the components of $\mathbf{q}_{S(N+3)}$. Supposing that the feasible region is not empty, nonlinear-programming algorithms will yield finite extreme value solutions to these problems if and only if the coherence of $P(X_{N+2})$ with $\mathbf{A}_{K,N}P(\mathbf{X}_N) \leq \mathbf{b}_K$ requires that $P(X_{N+2}) > 0$.

Proof of Part II. This result hinges on the coherency requirement that for any quantity X and event E, the assertion $P(X|E)$ must satisfy the restriction that $P(XE) = P(X|E)P(E)$. Thus, the coherency of an assertion $P(X_{N+1}|X_{N+2})$ with the assertions $\mathbf{A}_{K,N}P(\mathbf{X}_N) \leq \mathbf{b}_K$ requires that there be a vector $\mathbf{q}$ satisfying the linear restrictions specified in the theorem, for which $P(X_{N+2}) = \mathbf{r}_{N+2}\mathbf{q}$ and $P(X_{N+3}) = \mathbf{r}_{N+3}\mathbf{q}$, and for which $P(X_{N+1}|X_{N+2})\ \mathbf{r}_{N+2}\mathbf{q} = \mathbf{r}_{N+3}\mathbf{q}$. Thus, the relevance of the non-linear objective function $\mathbf{r}_{N+3}\mathbf{q}/\mathbf{r}_{N+2}\mathbf{q}$ to the coherence of the further assertion $P(X_{N+1}|X_{N+2})$ is established as long as $\mathbf{r}_{N+2}\mathbf{q}$ is bounded away from 0. However, no bound can be computed by these means for the quotient $\mathbf{r}_{N+3}\mathbf{q}_{S(N+3)}/\mathbf{r}_{N+2}\mathbf{q}_{S(N+3)}$ as long as there is a vector $\mathbf{q}$ satisfying the restrictions, for which $\mathbf{r}_{N+2}\mathbf{q} = 0$. This condition would allow an assertion of $P(X_{N+2}) = 0$ to cohere with $\mathbf{A}_{K,N}P(\mathbf{X}_N) \leq \mathbf{b}_K$. On the other hand, if all feasible vectors $\mathbf{q}$ entail that $\mathbf{r}_{N+2}\mathbf{q} > 0$, the quotient $\mathbf{c}_{N+3}\mathbf{q}/\mathbf{c}_{N+2}\mathbf{q}$ is necessarily bounded, for the feasible set of vectors $\mathbf{q}$ is closed and bounded. ∇

Figure 4.1 displays the logic of the argument. The numerical value of $P(X_{N+1}|X_{N+2})$ coherent with the assertions of $P(X_{N+2})$ and $P(X_{N+1}X_{N+2})$ equals the slope of the vector $(P(X_{N+2}),P(X_{N+3}))$ whenever a unique slope is defined. Suppose that the convex hull of the realm $\mathfrak{R}(X_{N+2},X_{N+3})$ is the dark bordered triangle with vertices (0,0),(1,1), and (1,5). If the coherency of $[P(X_{N+2}),P(X_{N+3})]$ with the assertion $\mathbf{A}_{K,N}P(\mathbf{X}_N) \leq \mathbf{b}_K$ requires that the vector be bound within the inscribed quadrilateral, for example, then it would also restrict a cohering assertion of conditional prevision $P(X_{N+1}|X_{N+2})$ to lie betwen the minimum and maximum slopes of lines through the origin that intersect the quadrilateral. If the assertion $(P(X_{N+2}),P(X_{N+3})) = (0,0)$ would cohere with $\mathbf{A}_{K,N}P(\mathbf{X}_N) \leq \mathbf{b}_K$, then every line through the origin would intersect the region of cohering assertions $(P(X_{N+2}),P(X_{N+3}))$, and thus there would be no bound on a cohering assertion of $P(X_{N+1}|X_{N+2})$ without strengthening the definition of coherency. Further discussion of this eventuality is beyond the scope of this paper.

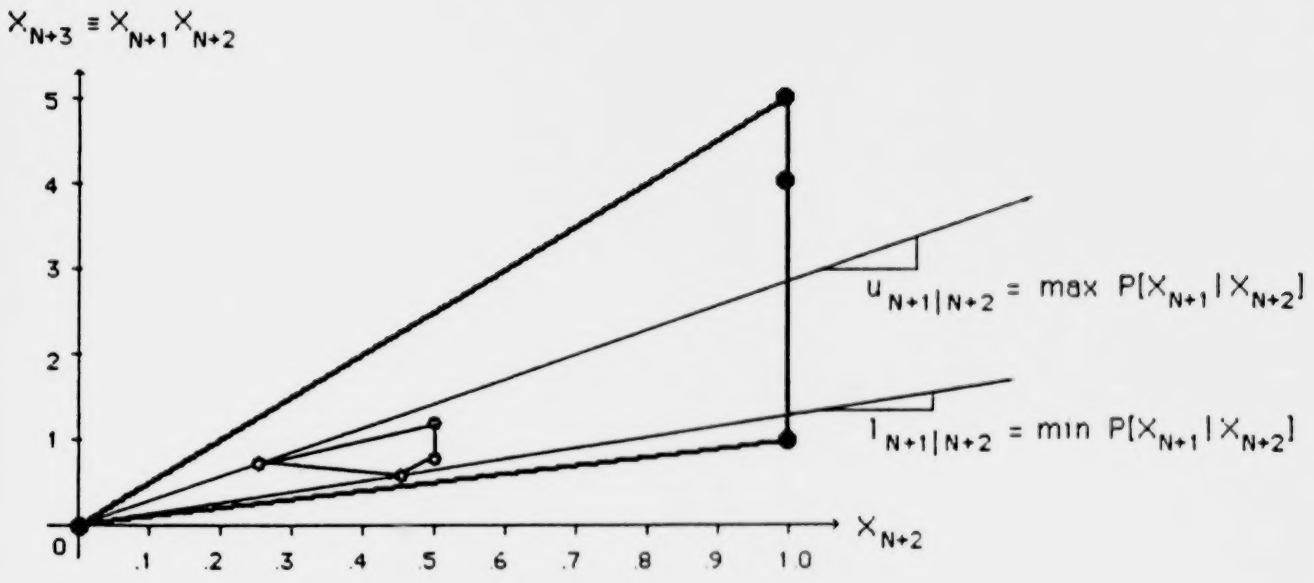

Figure 4.1. Coherence of conditional prevision. Bold dots represent vectors in $\mathcal{R}(X_{N+2}, X_{N+3})$, where $X_{N+3} \equiv X_{N+1}X_{N+2}$. The bold triangle is the boundary of their convex hull. The inscribed quadrilateral represents assertions of $P(X_{N+2}, X_{N+3})$ presumed to cohere with the assertions $\mathbf{A}_{K,N}P(\mathbf{X}_N) \leq \mathbf{b}_K$. Then the slopes denoted $l_{N+1|N+2}$ and $u_{N+1|N+2}$ are the minimum and maximum values of $P(X_{N+1}|X_{N+2})$ that would cohere with this partial assertion of $P(\mathbf{X}_N)$.

It is worth noting that Suppes (1981, p. 24) decried the non-existence of a result such as our FTP Part II as a "serious difficulty" for a numerical representation of uncertain knowledge.

A simple algorithm. The nonlinear programming problem of FTP Part II can be solved computationally by a one-dimensional monotonic search among solutions to related linear programming problems, as follows. To maximize the ratio y/x over points (x,y) of a closed polygon in the open right-half plane, define the linear function $z_\lambda(x,y) = y - \lambda x$, parameterized by λ. The equation $z_\lambda(x,y) = c$ represents a straight line of slope λ and the line passes through the origin $(0,0)$ only if $c = 0$. For given λ, we maximize the function $z_\lambda(x,y)$ over allowable (x,y), and write $\max z_\lambda(x,y) = z_\lambda(x_\lambda, y_\lambda) = c_\lambda$. (The line $z_\lambda(x,y) = c_\lambda$ now touches the allowable polygon only on its boundary, including

the point (x_λ, y_λ).) If a search is conducted and a value λ is found for which $c_\lambda = 0$, then

$$y_\lambda / x_\lambda = \lambda .$$

This slope is maximal since the line passes through the origin and touches the polygon only on its boundary. (For the minimization problem, use the same algorithm, with maximization replaced by minimization, throughout.)

Let us conclude this section by noticing that the strongest possible forms of two important inequalities are corollary to the Fundamental Theorem of Prevision.

Corollary 1. Completion of the Bienayme-Chebyshev Inequality. Let X be any quantity with finite discrete realm $\mathcal{R}(X) = \{x_1, x_2, \ldots, x_S\}$. Correspondingly, X^2 is a quantity with realm $\mathcal{R}(X^2) \equiv \{y \mid y = x^2, \text{ and } x \in \mathcal{R}(X)\} = \{x_1^2, x_2^2, \ldots, x_S^2\}$. Each event of the form $(X = x_i)$ is equivalent to the associated event $(X^2 = x_i^2)$; and all events of the form $(X = x_i)$, where $x_i \in \mathcal{R}(X)$, together constitute a partition. Denoting the vector of these constituents by $\mathbf{C}_S \equiv [(X = x_1), \ldots, (X = x_S)]^T$, we can write

$$\begin{bmatrix} X \\ X^2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_S \\ x_1^2 & x_2^2 & \cdots & x_S^2 \end{bmatrix} \mathbf{C}_S = \mathbf{\Pi}_{2,S}\, \mathbf{C}_S .$$

Suppose you assert precise numerical values for $P(X)$ and $P(X^2)$. Your variance for X is defined as $V(X) \equiv P[X - P(X)]^2 = P(X^2) - [P(X)]^2$, the latter equality being an implication of the coherency of your prevision. Now for any $\epsilon > 0$, define the event E_ϵ as the event that X differs from your $P(X)$ by at least ϵ, $E_\epsilon \equiv (|X - P(X)| \geq \epsilon)$. Finally, let $\mathbf{r}_\epsilon$ denote that row vector with components 0 or 1 for which $E_\epsilon = \mathbf{r}_\epsilon \mathbf{C}_S$. Then for any $\epsilon > 0$, coherency requires that your $P(E_\epsilon)$ lie within the interval bounded by the solutions to the following two linear programming problems:

Find the vectors $\mathbf{q}_S$ that minimize and maximize $\mathbf{r}_\epsilon \mathbf{q}_S$

subject to the restrictions that

$$\begin{bmatrix} P(X) \\ P(X^2) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_S \\ x_1^2 & x_2^2 & \cdots & x_S^2 \end{bmatrix} \mathbf{q}_S$$

that $\Sigma \mathbf{q}_S = 1$, and that each component of the vector $\mathbf{q}_S$ be non-negative. Δ

The familiar Bienaymé-Chebyshev inequality states the weaker conclusion that under the conditions specified, $P(E_\epsilon) \leq V(X)/\epsilon^2$. The major use of that inequality has been in proving various forms of the weak law of large numbers. For in practice, the traditional statement of the upper bound is notoriously large, often too large to be useful. The inequality stated above as corollary to the fundamental theorem of prevision actually strengthens the inequality as proved by Chebyshev (1867) to the most extreme statement that can be made in any particular application. A computational example showing such an improvement appears in Section 5. Moreover, our corollary completes the celebrated inequality by specifying a lower bound as well as an upper bound on your prevision for the event $(|X - P(X)| \geq \epsilon)$ in any given instance.

As mentioned, the Bienaymé-Chebyshev inequality has found its widest use in theoretical studies of the weak law of large numbers. The weak law concerns bounds on your probability that the average of several quantities deviates from your prevision for the average by more than any specified amount: $P(|\overline{X}_N - P(\overline{X}_N)| \geq \epsilon)$. Obviously, the Bienaymé-Chebyshev inequality is relevant if you assert your $P(\overline{X}_N)$ and $P[(\overline{X}_N)^2]$. The strong law of large numbers concerns bounds on the less restrictive event that at least one member of a sequence of averages so deviates: $P(\text{Max}_{0 \leq k \leq K} |\overline{X}_{M+k} - P(\overline{X}_{M+k})| \geq \epsilon)$ for specified values of M and K. The fundamental theorem of prevision provides as corollary a necessary and sufficient bound for coherent previsions of such extreme events. We state this corollary in the context of any finite sequence of discrete quantities that you regard as exchangeable, the paradigmatic context for statistical inference. Our corollary differs from the usual Kolmogorov inequality, first, in assuming exchangeability instead of independence and, secondly, by involving successive averages directly, rather than sums of quantities.

Let us denote by $\mathbf{X}_N$ the vector of quantities $(X_1, \ldots, X_N)^T$, having a common realm $\mathcal{R}(X_j) = \mathcal{R}(X) = \{x_1, \ldots, x_S\}$. (So S denotes the size of this common realm of the components.) Similarly, $\mathbf{X^2}_N$ denotes the corresponding vector of the squares of these quantities, and $(\mathbf{X_i X_j})_N$ denotes the vector of the $N(N-1)/2$ product quantities $X_i X_j$, where $1 \leq i < j \leq N$. Finally, we denote by $\mathbf{C}_{S(N)}$ the vector of constituents of the partition composed of the events of the form

$(\mathbf{X}_N = \mathbf{x}_N)$, one for each element $\mathbf{x}_N$ in the realm $\mathcal{R}(\mathbf{X}_N)$. Using this notation, we write

$$\begin{bmatrix} \mathbf{X}_N \\ \mathbf{X^2}_N \\ (\mathbf{X_i X_j})_N \end{bmatrix} = \begin{bmatrix} \mathbf{R}(\mathbf{X}_N) \\ \mathbf{R}(\mathbf{X^2}_N) \\ \mathbf{R}((\mathbf{X_i X_j})_N) \end{bmatrix} \mathbf{C}_{S(N)} \quad .$$

Each submatrix $\mathbf{R}(\cdot)$ is composed of columns that are the appropriate vector members of the realm of the quantity vector shown within the parentheses. Notice that S(N) may be any positive integer between S and S^N, depending on the logical relations embedded in the definitions of the components of $\mathbf{X}_N$.

Corollary 2. Bounds on Probabilities of Extreme Sequences. Let $X_1, \ldots, X_N$ be a sequence of quantities which you regard as exchangeable. Suppose you assert three precise numbers for your $P(X_1) = P(X_i)$, for all i $(1 \le i \le N)$, your $P(X_1^2) = P(X_i^2)$ $(1 \le i \le N)$, and your $P(X_1X_2) = P(X_iX_j)$ $(1 \le i < j \le N)$. For each positive integer M and each non-negative integer K, and for any $\epsilon > 0$, define the event $E_{M,K,\epsilon} \equiv (\text{Max}_{0 \le k \le K} |\overline{X}_{M+k} - P(\overline{X}_{M+k})| \ge \epsilon)$, where $\overline{X}_T$ denotes the arithmetic average of the quantities $X_1, \ldots, X_T$. Your presumed assertion of exchangeability requires that your $P(\overline{X}_{M+k}) = P(X_1)$. Finally, let $\mathbf{r}_{M,K,\epsilon}$ be the indicator row vector for which $\mathbf{r}_{M,K,\epsilon}\mathbf{C}_{S(N)} = E_{M,K,\epsilon}$. Then coherency requires that your $P(E_{M,K,\epsilon})$ lie within the interval bounded by the extreme values of $\mathbf{r}_{M,K,\epsilon}\mathbf{q}_{S(N)}$, subject to the appropriate linear restrictions generated from your assertions,

$$\begin{aligned} P(\mathbf{X}_N) &= P(X_1)\mathbf{1}_N = \mathbf{R}(\mathbf{X}_N)\mathbf{q}_{S(N)} \quad , \\ P(\mathbf{X^2}_N) &= P(X_1^2)\mathbf{1}_N = \mathbf{R}(\mathbf{X^2}_N)\mathbf{q}_{S(N)} \quad , \text{ and} \\ P[(\mathbf{X_i X_j})_N] &= P(X_1X_2)\mathbf{1}_{N(N-1)/2} = \mathbf{R}((\mathbf{X_i X_j})_N)\,\mathbf{q}_{S(N)} \quad . \end{aligned}$$

along with $\Sigma\mathbf{q}_{S(N)} = 1$ and all components of $\mathbf{q}_{S(N)}$ non-negative. Moreover, your regarding the quantities X_i as exchangeable places additional linear requirements on the vector $\mathbf{q}_{S(N)}$. Any components of $\mathbf{q}_{S(N)}$ must be equal if the corresponding columns of the realm matrix $\mathbf{R}(\mathbf{X}_N)$ are permutations of one another. These exchangeability restrictions can be expressed in the form $\mathbf{M}\mathbf{q}_{S(N)} = 0$, where each row of the matrix $\mathbf{M}$ contains one 1, one -1, and 0 in the remaining S(N)-2 positions. Δ

Although only corollary to our fundamental theorem of prevision, this is a very general and strong statement of its own. It specifies the strictest bounds on your prevision (probability) for events of the form $E_{M,K,\epsilon}$ that are implied by your avowed assessment of N quantities as exchangeable and your prevision assertions as stated in the theorem. (Kolmogorov's inequality and the usual statement of the strong law of large numbers presuppose your stronger assertion of independence regarding $X_1, \ldots, X_N$. Such an assertion would be representable by a further specification of polynomial restrictions on the components of $\mathbf{q}_{S(N)}$ that we will not describe here in detail.) Within the minimalist conception of mathematics subscribed to by de Finetti, the laws of large numbers are well specified properties of prevision for events of the form $E_{M,K,\epsilon}$, where M and K have specific finite values. Detailed discussion appears in several sections of de Finetti (1970: I, 6.8; II, 7.5). Our Corollary 2 to the fundamental theorem of prevision states precisely the sharpest bounds on a prevision $P(E_{M,K,\epsilon})$ that are necessary and sufficient for its coherence with the asserted previsions mentioned. Thus, the corollary identifies the asserted status of any sequence of quantities vis-a-vis the law of large numbers condition, $P(\text{Max}_{0 \le k \le K} |\overline{X}_{M+k} - P(\overline{X}_{M+k})| \ge \epsilon) \le \delta$ for specified values of M and K, that is required by its coherency with the assertion of exchangeability regarding component quantities. More standard specifications of the status of exchangeable sequences in terms of limit theorems are compiled in the monograph of Taylor, Doffer, and Patterson (1985). It is somewhat ironic that such a simple characterization of coherent probabilities relevant to the laws of large numbers is achieved within the operational-subjective formulation of probability via the fundamental theorem of prevision. For the laws of large numbers, so central to objectivist theories of probability such as the frequency theory and the propensity theory, are only a curiosity in the subjectivist theory, which centers upon practical questions of your knowledge about particular finite sequences.

In the next section, we present small computational examples and suggest realistic applications of our arguments.

5. Computations and Applications

To begin, we illustrate our improvement on the Bienaymé-Chebyshev inequality, which pertains to a single quantity. We shall then extend the context to several quantities regarded as exchangeable, in order to illustrate the computable bounds on coherent probabilities of extreme sequences.

Example 1. Suppose that X is a quantity with realm $\mathcal{R}(X) = \{1,2,3,4,5\}$. Thus, we can write

$$\begin{bmatrix} X \\ X^2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 9 & 16 & 25 \end{bmatrix} \mathbf{C}_5 = \mathbf{R}_{2,5}\,\mathbf{C}_5 \quad ,$$

where $\mathbf{C}_5$ is the column vector of events $[(X=1),(X=2),(X=3),(X=4),(X=5)]^T$. The convex hull of the realm $\mathcal{R}[(X,X^2)^T]$ is the dark bordered polygonal region depicted in Figure 5.1. To begin this example, suppose you assert the previsions $P(X) = 2.2$ and $P(X^2) = 7$, or equivalently, your $V(X) = P(X^2) - [P(X)]^2 = 2.16$. The point $[P(X), P(X^2)] = [2.2, 7.0]$ should be identifiable in the figure. The figure also shows that the assertion of $P(X^2)$ within the interval $[5.0, 8.2]$ is necessary and sufficient for its coherence with the assertion $P(X) = 2.2$. In the course of this extended example, we will also consider alternative assertions, $P(X^2) = 6.0$ and $P(X^2) = 7.6$.

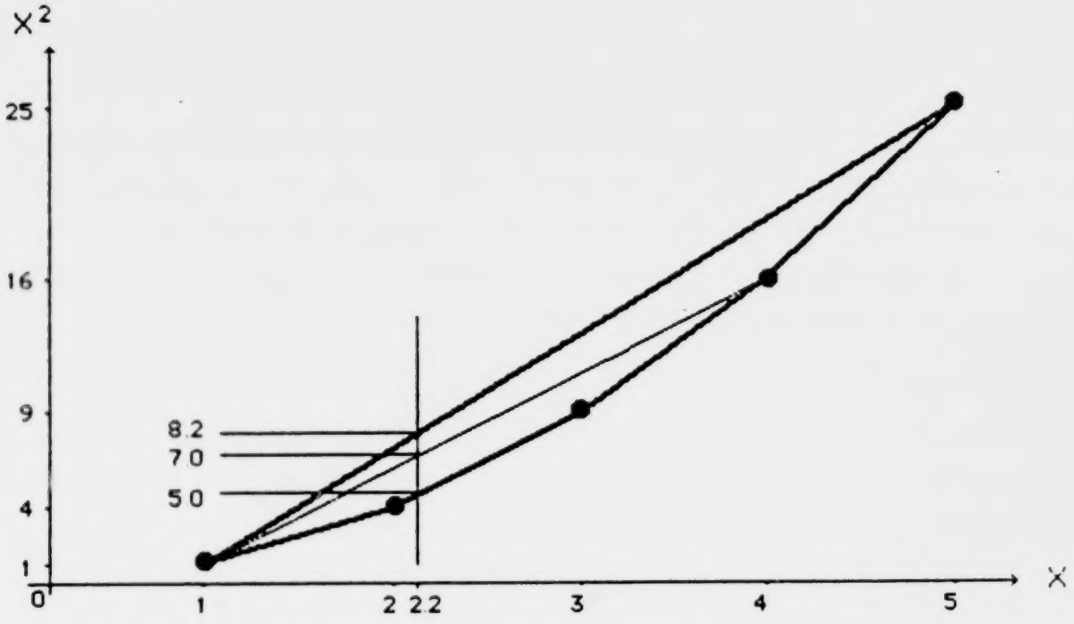

Figure 5.1. Completion of Bienaymé-Chebyshev inequality. The convex hull of $\mathcal{R}(X, X^2)$ is divided into two regions by a diagonal line. The lower part, including the points connected by the dividing line, is the convex hull of the subset of possible observations for which $|X - 2.2| < \epsilon$, provided ϵ satisfies the inequality $1.8 < \epsilon \leq 2.8$.

Let us first study an event that is easy to describe geometrically: the event that X differs from 2.2, your P(X), by at least 2.8 units. Using the notation of Corollary 1, we write $E_{2.8} = (|X - 2.2| \geq 2.8) = (0\,0\,0\,0\,1)\,\mathbf{C}_5$. What

does the coherency of your prevision require of your $P(E_{2.8})$? Since $E_{2.8}$ is a function of X, you can imagine a third axis for this quantity, rising perpendicularly up out of the plane of Figure 5.1. Visualizing the 3-dimensional figure, we see the value $E_{2.8} = 1$ if X has the value 5. But for the other four possible values of X, $E_{2.8} = 0$. Now the convex hull depicted in the original plane figure can be viewed as the projection of the convex hull of $\mathcal{R}(X, X^2, E_{2.8})$ onto the space of (X, X^2). The lower half polygon whose vertices are the four points (1,1,0), (2,4,0), (3,9,0), and (4,16,0) constitutes the bottom face of the 3-dimensional hull. There are four other faces on this hull. Each is defined by a triangle connecting one edge of this bottom face with the point (5,25,1).

Since any coherent prevision point for the vector of quantities $(X, X^2, E_{2.8})$ must lie within the hull in three dimensions, it should be evident why the linear programming solution that minimizes $P(E_{2.8})$ subject to the relevant restrictions yields a lower bound of 0 (corresponding to the primal solution vector $\mathbf{q}_5 = (.6, 0, 0, .4, 0)$). For the associated prevision vector $P(X, X^2, E_{2.8}) = (2, 7.0, 0)$ lies on the bottom face of the hull. The maximization problem subject to the same constraints yields an upper bound of .2 (corresponding to the primal solution vector $\mathbf{q}_5 = (.4, .4, 0, 0, .2)$). The associated prevision vector $P(X, X^2, E_{2.8}) = (2, 7.0, .2)$ is the highest point in the hull that projects onto (2.2, 7.0) [in the space of (X, X^2)]. This upper bound on $P(|X - 2.2| \geq 2.8)$ is sharper than the Bienaymé-Chebyshev bound in this case: $V(X)/\epsilon^2 = 2.16/(2.8)^2 \approx .276$. Notice that Figure 5.1 would be unchanged for illustrating the logic of coherent prevision for any other event E_ϵ for which ϵ lies within the half-open interval (1.8, 2.8]. Events such as $E_{1.81} \equiv (|X - P(X)| \geq 1.81)$ and $E_{2.8} \equiv (|X - P(X)| \geq 2.8)$ are identically equivalent to the event $(X = 5)$.

Figure 5.1 can also be used to aid ones intuition in several more of the computational results presented below. Considering a coherent prevision for the event $E_{1.2} = (|X - 2.2| \geq 1.2)$, restricted only by the assertions $P(X) = 2.2$ and $P(X^2) = 7$, one recognizes that the triangle connecting the points (1,1,1), (4,16,1), and (5,25,1) constitutes the upper face of the convex hull of $\mathcal{R}(X, X^2, E_{1.2})$, while the line connecting (2,4,0) with (3,9,0) constitutes a lower edge. It should then be evident that the upper bound for coherent $P(E_{1.2})$ is 1, while the lower bound will exceed 0. The precise lower bound is .6, as listed with the upper and lower bounds for various values of ϵ and V(X) in Table 5.1, below. As shown for

$\epsilon = 1.2$, if you assert $P(X) = 2.2$ and $P(X^2) = 6$, coherency of prevision will bound your assertion of $P(E_{1.2})$ within an interval $[l,u]$ that lies strictly within $(0,1)$. The relevant upper bound computed from the inequality of Bienaymé-Chebyshev appears following our computed interval in each case.

Reading down the columns of Table 5.1, notice that both the upper bound and the lower bound decrease (weakly) as ϵ increases. But reading across a row, say when $\epsilon = 1.8$, notice that there is not a monotonic pattern in the upper bound on $P(E_{1.8})$ with increasing values of $P(X^2)$. This latter result may appear counter-intuitive to readers unduly influenced by their experience with the Bienaymé-Chebyshev inequality. If one makes the appropriate adjustments in visualizing Figure 5.1 to illustrate $E_{1.8}$, one will see the interesting reason why the upper bound for coherent $P(E_{1.8})$ is smaller when $P(X^2) = 7.6$ than it is when $P(X^2) = 7.0$.

Table 5.1 Bounds* on prevision for Chebyshev's event, $P(E_\epsilon)$, necessary and sufficient for its coherence with various specified values of $P(X)$ and $P(X^2)$, followed by the Bienaymé-Chebyshev upper bound, headed u_{B-C}.

$P(X, X^2)$	(2.2, 6.0)		(2.2, 7.0)		(2.2, 7.6)	
ϵ	coherent bounds	u_{B-C}	coherent bounds	u_{B-C}	coherent bounds	u_{B-C}
0.8	[.267, 1.0]	1.813	[.6 , 1.0]	3.375	[.8 , 1.0]	4.313
1.2	[.267, .5]	.806	[.6 , 1.0]	1.5	[.8 , 1.0]	1.917
1.8	[.025, .233]	.358	[.15, .4]	.667	[.225, .35]	.852
2.8	[0.0 , .117]	.148	[0.0 , .2]	.276	[.15 , .25]	.352

*The bounds presented are accurate to the nearest one-thousandth.

Example 2. Expanding consideration to several quantities, we provide an example illustrating Corollary 2, which specifies bounds on coherent probabilities for extreme events. Suppose that X_1, X_2, and X_3 are logically independent quantities with the common realm $\mathcal{R}(X_i) = \{1,2,3,4,5\}$, as

in Example 1. Table 5.2 exhibits the computed upper and lower bounds on previsions for events of the form $E_{M,K,\epsilon} \equiv (\text{Max}_{0 \le k \le K} |\bar{X}_{M+k} - P(\bar{X}_{M+k})| \ge \epsilon)$ that would cohere with three different assertion configurations regarding the quantities X_1, X_2, and X_3. Along any row that begins with a specification of M,K,ϵ, appear the intervals for $P(E_{M,K,\epsilon})$ coherent with the mere assertions for all i,j of $P(X_i) = 2.2$ and $P(X_i^2) = 6.0$, along with $P(X_iX_j)$ appropriate to characterize the specified correlation $\rho(X_i,X_j)$ that heads each column. [When $P(X_i,X_i^2) = (2.2,6.0)$, the correlations $\rho(X_i,X_j)$ equal to 0, .25, and .75 are implied, respectively, by the additional assertion of $P(X_iX_j)$ equal to 4.84, 5.13, and 5.71.] In the subsequent row are printed the lower and/or upper bound in any case for which the coherent bounding interval is restricted further by the additional assertion of exchangeability regarding X_1, X_2, and X_3. Notice that the additional restriction shrinks the interval further whenever it has any effect.

Table 5.2. Bounds* on $P(E_{M,k,\epsilon})$ necessary and sufficient for its coherence with $P(X_i,X_i^2) = (2.2,6.0)$ and specified values of $\rho(X_i,X_j)$, without and with the assertion of exchangeability for X_1, X_2, and X_3.

$\rho(X_i,X_j)$	0.0	.25	.75
M,K,ϵ			
1,1,0.8	[.267 , 1.0]	[.267 , 1.0]	[.267 , 1.0]
1,1,1.8	[.025 , .311]	[.025 , .311]	[.025 , .288]
	.026 .233	.233	.233
1,2,0.8	[.267 , 1.0]	[.267 , 1.0]	[.267 , 1.0]
1,2,1.8	[0 , .333]	[.025 , .345]	[.025 , .318]
	.026 .233	.233	.233
2,1,0.8	[.042 , .920]	[.059 , 1.0]	[.101 , 1.0]
	.088 , .788]	.068 .974	
2,1,1.8	[0 , .175]	[0 , .220]	[0 , .258]
	.121	.158	.230

*The bounds presented are accurate to the nearest one-thousandth.

Example 3. This final example presents computational results illustrating implications of Part II of the Fundamental Theorem of Prevision for a further *conditional* prevision. The context for this example continues from Example 1. Remember that $\mathcal{R}(X) = \{1,2,3,4,5\}$, and $E_\epsilon \equiv (|X - P(X)| \geq \epsilon)$. Each column of Table 5.3 is headed by vector values for an assertion of $P(X,X^2)$. Each row of the table identifies a specific event of the form E_ϵ. In the intersection of each row and column appear the bounds for a further assertion of $P(X|E_\epsilon)$ if it is to cohere with the assertion identified by the column heading. These bounds were computed via the nonlinear programming problems identified in our Fundamental Theorem of Prevision, Part II. [Notice by the earlier Table 5.1, that the prevision for each of the events E_ϵ listed in Table 5.3 is bounded away from 0 by the requirement that it cohere with the assertions of $P(X,X^2)$.] Figure 4.1, which appeared in the previous section to illustrate the proof of this part of our theorem, is drawn to a scale that illustrates this example under the specifications $P(X,X^2) = (2.2,6.0)$ and $\epsilon = 1.2$.

Table 5.3 Bounds* on a conditional prevision $P(X|E_\epsilon)$ necessary and sufficient for its coherence with various specified values of $P(X)$ and $P(X^2)$.

$P(X,X^2)$ / E_ϵ	(2.2, 6.0)	(2.2, 7.0)	(2.2, 7.6)
$(\|X-2.2\| \geq 0.8)$	[2.200 , 2.750]	[2.200 , 2.333]	[2.200 , 2.250]
$(\|X-2.2\| \geq 1.2)$	[1.222 , 2.750]	[1.857 , 2.333]	[2.059 , 2.250]
$(\|X-2.2\| \geq 1.8)$	[4.000 , 5.000]	[4.000 , 5.000]	[4.429 , 5.000]

*The bounds presented are accurate to the nearest one-thousandth.

Moving beyond these simple computational illustrations, we suggest by example the vast potential for practical applications. Complex engineering systems such as nuclear power plants or space vehicles are made up of many component subsystems, with various dependencies between components, some providing backups for others via intricate linkages. Typically, the operating status of the overall system can be represented as a complicated logical

function of the status of many components. Yet quality testing in the design and construction of the system usually can be conducted only on a component-by-component basis. After such testing, engineers may be able to assert their previsions for the status of individual components under various conditions, and perhaps even for a few of such components in conjunction. But it may be difficult for anyone, and even for a team, to assess directly the operating status of the system as a whole. The linear programming method underlying the fundamental theorem of prevision can be used to keep a running track of the bounds on coherent prevision for the status of the system implied by the changing assertions of engineers concerning the status of components.

A more standard statistical application involves conditional prevision assertions regarding characteristics of a finite population of which some subgroup has been observed. One application with which we are familiar concerns the annual milk yields of a group of 27 thousand dairy cows whose yields are regarded as exchangeable by a dairy expert. An exact yield has been recorded for some 850 of these cows. Specific assertions made by the expert about the yields from cows of this type can be inserted as input in the programming problems to determine the bounds on cohering conditional prevision assertions about the unobserved yields given the observed yields.

The practitioner may well react with horror at the huge computational dimension of the programming problems that could be involved in realistic applications. (Annual yields from individual cows of this particular type can range realistically from 12 thousand pounds to 40 thousand pounds, so even the realm of each observation can be immense, depending on the fineness of resolution in the reported yield.) Two quieting remarks are in order. First, without elaboration here, let us mention that large reductions in the dimension of the programming problems can be achieved algebraically by making more efficient computational use of the exchangeability which has been specified. In example 2 discussed above, the dimension of the activity vector in our actual computations was reduced from 125 to 35. Secondly, the computational time for solving large linear programming problems is reduced from exponential to polynomial time by the ellipsoid methods of Shor and Khachian (1979), and more recently, Karmarkar (1984). The survey article by Bland, Goldfarb, and Todd (1981) and the textbook introduction of Walsh (1985) are helpful. Coupling these with the benefits of simultaneous processing achieved by supercomputers, or banks of microcomputers, we feel that even realistically large scale problems could be accessible to computation.

In this tentative happy mood, let us comment on the applicability of the fundamental theorem of prevision to another sizeable practical problem. Economists at several institutions regularly produce quarterly forecasts for macroeconomic measurements of the U.S. economy. Brayton and Mauskopf (1985) described a recent version of the Federal Reserve Board forecasting model containing some 332 equations and 124 forecast variables. In conventional statistical terminology, it is recognized that the large size of such models, and their many lags and nonlinearities, preclude the application of simultaneous estimation techniques. Thus, the many equations are usually estimated singly. Litterman (1986) and McNees (1986) each noted that forecasters' subjective judgments are typically appended to model-based computations to produce a useful forecast. These judgments are based on both an analysis of residuals from individual equations and on intermediate monthly observations of those components of the quarterly statistics that are also recorded monthly. Moreover, applied economists who are knowledgeable of even daily information on particular sectors such as housing construction, inventories, capital investment, or capacity utilization can provide a wealth of relevant information which is not amenable to systematic recording in a prior-formatted data file. How are all these sources of information to be incorporated into a coherent prevision assessment for quantities which are of interest for policy decisions? The fundamental theorem of prevision provides a computational framework in which judgments based on a variety of information sources can be accumulated and their coherency checked.

6. Conclusions

We hope that the substantive statistical results of this paper will lead you to consider the Fundamental Theorem of Prevision deserving of its appellation. Our concluding discussion will run in a philosophical vein. We use both standard logical notation and the arithmetical notation for logical relations. The latter was established by Boole (1847) and was used by de Finetti (1967). In arithmetic notation, the sentence $(E_1 \wedge E_2)$ is expressed as the product E_1E_2, and the sentence $\sim E$ is expressed as $(1-E)$. Thus, for example, the sentence "E_1 implies E_2", written in logical notation as $\sim[E_1 \wedge \sim E_2]$, is expressed arithmetically as $\{1 - [E_1(1-E_2)]\}$, or $(1 - E_1 + E_1E_2)$. Such a sentence can be true or false (the arithmetic quantity can equal 1 or 0) depending on the truth of the component propositions E_1 and E_2 (their numerical values).

That the syllogism is without content has long been a subject of logicians' musings. An important development in the programme of devising how the notion of "content" could be integrated into the formal expression of knowledge was Frege's (1879) distinction between a *contemplated* sentence, denoted by E, and an *asserted* proposition (by you, by someone), denoted by $\vdash E$. Ostensibly, Frege was no friend of the subjectivist stance. Known among statisticians for his ranting against "psychologism" in the field of logic (Frege, 1893), and perhaps for his provocation of Russell's paradox (Van Heijenoort, 1967), he is unfortunately less well known for the acumen of many of his ideas (Resnik, 1980). De Finetti, for example (1970, 2.6), mistakenly attributed the proposed distinction between contemplated and asserted propositions to Koopman (1940). And Jeffreys (1961, I,§1.51) noted only its use by Whitehead and Russell (1910). Levy (1980) contains insightful critical discussion.

De Finetti lauded the distinction, however, remarking that we should recognize prevision as an assertion (by you, by someone). But he declined to use the assertion notation, supposing that this distinction would be clear from the context. In two-valued, "deductive", logic, your asserting something about a sentence such as $(E_1 \Rightarrow E_2)$ may take only two possible forms: you may assert that the sentence is true, $\vdash (1 - E_1 + E_1E_2) = 1$; or you may assert that it is false, $\vdash (1 - E_1 + E_1E_2) = 0$. This is the rule of two-valued logic. In the many-valued logic of coherent prevision, your assertion can take the form $P(1 - E_1 + E_1E_2) = *$, where this number may be any number in the interval $[0,1]$. Thus, the symbol P replaces and expands the assertion symbol $\vdash$ of two-valued logic.

Frege's distinction allows you to contemplate a sentence without asserting either that it is true, or that it is false. Indeed, within the confines of two-valued logic, this is the only weakening possible from the full throated assertion that a sentence is true or that it is false. The syllogisms of deductive logic specify equivalence relations among well-formed-formulae within the logic. The considered formulae are equivalent irrespective of whether or not anyone asserts the sentences to be either true or false. A person's willingness to be understood in this logic is signified by accepting the logical law of noncontradiction, $E \wedge \tilde{E} \equiv E(1-E) = 0$, along with all its consequences, such as $E \vee \tilde{E} \equiv E + (1-E) - E(1-E) = 1$. Thus, within this logic, any assertion regarding a sentence E that is equivalent to the assertion $\vdash (E \vee \tilde{E}) = 1$ amounts to no

assertion at all about E. It is a redundancy relative to the person's presumed willingness to be understood in this logic. To be sure, no one can be *forced* to make an assertion about any sentence that is not determined by the principle of noncontradiction. You need neither assert E, $\vdash E = 1$, nor assert $\tilde{E}$, $\vdash E = 0$. In the extreme, you may find yourself in the non-assertive contemplative position: $\vdash (E \vee \tilde{E}) = 1$, an assertion without content. The principle of coherency is merely the extension of the principle of non-contradiction to the many-valued logic of uncertain knowledge. As in deductive logic, there is no compulsion that anyone make an assertion about any quantity. Just as $\vdash (E \vee \tilde{E}) = 1$ is a redundant "assertion" without content for anyone committed to communication within the confines of deductive logic, your partial assertion that your $P(X) \in [\min \mathcal{R}(X), \max \mathcal{R}(X)]$ is a redundant assertion without content in the logic of prevision. It amounts to no assertion at all if you accept the principle of coherency, which is necessary for communication within this logic.

Long a stumbling block to the acceptability of subjective Bayesian statistical procedures has been the objection "But for many quantities, I am in no state of mind to assert my prevision. I cannot now assert anything about **X**." Subjectivists have annoyedly responded, "Sure you can. It just takes effort on your part to elicit your prevision. Just try to do the best you can." Shafer (1976) has spiritedly and repeatedly suggested that the (non)assertion $P(X) \in [\min \mathcal{R}(X), \max \mathcal{R}(X)]$ is what represents one's knowledge (that is, lack of knowledge) in such instances. Both Shafer's insistence that probability bounds are not meant to represent betting odds, and his general proposed schema of inference have drawn appropriate criticism that his probability intervals have no operational meaning, and that his schema supports incoherent assessments. [See for instance the comments of Lindley, of Good, and of Hill in the discussion to Shafer (1982).] But Leamer's insight (1986) that a bounding statement, such as $P(X) \in [a,b]$, could be interpreted operationally as a "bid-ask spread" resolves the impasse to accepting Shafer's proposal in this instance. In the context of the Fundamental Theorem of Prevision, this appears to be a beautiful resolution to the search for a distribution that represents "ignorance", a search which unfortunately has intrigued many. There is no *distribution* that can represent uniquely the assertion of ignorance: $P(X) \in [\min \mathcal{R}(X), \max \mathcal{R}(X)]$. It is a prevision polytope identical to the convex hull of $\mathcal{R}(X)$ that represents this lack of knowledge, this lack of sufficient motivation to assert anything about X.

Their commitment to the operational-subjective formulation of probability notwithstanding, both Savage and de Finetti were disconcerted by the practical problem of identifying one's prevision *exactly* - though expressly no more than by the prospect of measuring anything *exactly* (de Finetti, 1970, Appendix 19.3). Savage recorded his qualifications already in 1954 (Ch. 4). Together (de Finetti and Savage, 1962) they wrote extensive commentary on the relevant article of Smith (1961). And de Finetti's final appendices (1970, Appendices 14-19) discuss the issues with his customary exhausting brilliance. One important insight which eluded their analysis was the understanding of an operational meaning to a partially asserted prevision polytope.

We submit that the complete extension of our understanding of representations for all forms of uncertain knowledge, no matter how rich nor how meagre, is provided by the Fundamental Theorem of Prevision. It supports the conclusions which we have expressed above. Coherent uncertain knowledge of a quantity vector is representable by a convex polytope within the convex hull of the realm of the quantity vector. Central to de Finetti's minimalist approach to mathematical construction was his rejection of a "preconceived preference for that which yields a unique and elegant answer *even when the exact answer should be 'any value lying between these limits'*." (1970, 6.3) The Fundamental Theorem of Prevision applies the framework of weak mathematical formulations to the characterization of states of uncertain knowledge by means of an asserted prevision polytope.

ACKNOWLEDGEMENTS

This paper summarizes joint research by the authors, supported in its initial phase by the State University of New York at Albany, Department of Mathematics and Statistics and Department of Economics. We have profited from comments on earlier versions of this work presented at the Universities of Trieste, Rome, and Hull, and at the Special Studies Research Section of the Federal Reserve Board of Governors in Washington, D.C. Work by Professor Dickey was supported by the U. S. National Science Foundation, Research Grant Number DMS-8614793.

References

BLAND, R.G., GOLDFARB, D., AND TODD, M.J. (1981). The ellipsoid method: a survey. *Operations Research*, **29**, 6, 1039-1091.

BOOLE, G. (1847). *The Mathematical Analysis of Logic*. London: George Bell.

BOREL, E. (1924). Apropos of a Treatise on Probability, *Revue Philosophique*. English translation by H. Smokler in H. Kyburg and H. Smokler (eds.), *Studies in Subjective Probability*, New York: John Wiley, 1964.

BRAYTON, R. AND MAUSKOPF, E. (1985). The M P S Model of the United States Economy . *Economic Modelling*, **2**(3), 170-292.

BRUNO, G. AND GILIO, A. (1980). Applicazione del metodo del simplesso al teorema
fondamentale per le probabilita nella concezione soggettivistica. *Statistica*, **40**, 3, 337-344.

CHEBYSHEV, P. L. (1867). Des valeurs moyennes. *Journal de Mathématiques Pures et Appliquées*, **12**, 177-184. English translation by H. M. Walker in D. E. Smith (ed.), *A Sourcebook of Mathematics*, Vol. II, New York: Dover.

CHERNOFF, H. (1961). A new approach to the evaluation of multiple choice questionnaires. *Bulletin of the International Statistical Institute*, **39**, 4, 289-293.

_____ (1962). The scoring of multiple choice questionnaires. *Annals of Mathematical Statistics*, **33**, 375-393.

COOMBS, C.H., MILHOLLAND, J.E., and WOMER, F.B. (1956). *Educational Psychology and Measurement*, **16**, 13-37.

DELL'ERA (1964). Scoring systems for the evaluation of multiple-choice tests. Unpublished Ph.D. thesis, University of Rome.

DEMPSTER, A. (1967). Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics*, **38**, 325-339.

DE FINETTI, B. (1937). La Prevision: ses lois logiques, ses sources subjective. *Annales de l'Institut Henri Poincare*, **7**. English translation by H. Kyburg in H. Kyburg and H. Smokler, eds., *Studies in Subjective Probability*, New York: Krieger, 1980, Second edition is free of typographical errors.

_____ (1949). Sull impostazione assiomatica del calcolo delle probabilita. *Annali Triestini dell'Universita di Trieste*, ***XIX***, 29-81. English translation by G. Majone in B. de Finetti, *Probability, Induction, and Statistics*, New York: John Wiley, 1972.

_____ (1965). Methods for discriminating partial knowledge concerning a test item, *British Journal for Mathematical and Statistical Psychology*, **18**, 87-123.

_____ (1967). Quelques conventions qui semblent utiles. *Revue Romaine des Mathematiques Pures et Appliques*, **12**, 1227-1233. English translation by L. J. Savage in B. de Finetti, *Probability, Induction, and Statistics*. New York: John Wiley, 1972.

_____ (1970). *Teoria delle Probabilita*. Torino: Einaudi. English translation by A. Machi and A.F.M. Smith, *Theory of Probability*. 2 Volumes, New York: John Wiley, 1974, 1975.

DE FINETTI, B. AND SAVAGE, L. J. (1962). Sul modo di scegliere le probabilita' iniziali. Reprinted in *The Writings of Leonard J. Savage*, Washington D.C.: American Statistical Association, 1981.

FISHBURN, P. (1965). Analysis of decisions with incomplete knowledge of probabilities. *Operations Research*, **13**(2), 217-237.

_____ (1985). *Interval Orders and Interval Graphs: A Study of Partially Ordered Sets*. New York: John Wiley Interscience.

FREGE, G. (1879). Begriffschrift, a formula language, modeled on that of arithmetic, for pure thought. English translation by S. Bauer-Mengelberg, printed in Van Heijenoort (1967).

_____ (1893). *The Basic Laws of Arithmetic: Exposition of the System*. English translation by M. Furth. Berkeley: University of California Press, 1967.

GOLDSTEIN, M. (1985). Temporal coherence, *Bayesian Statistics 2*. Edited by J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith. Amstrdam: Elsevier Science; and Valencia University Press.

GOOD, I. J. (1950). *Probability and the Weighing of Evidence*. New York: Hafner.

HILBERT, D. and ACKERMANN, W. (1938). *Grundzuge der Theoretischen Logik*, Berlin: Springer. English translation by L. Hammond, G. Leckie, and F. Steinhardt, *Mathematical Logic*, New York: Chelsea, 1950.

JEFFREYS, H. (1961). *Theory of Probability*, Third Edition, Oxford: Clarendon.

KARMARKAR, N. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica* **4** (4) 373-395.

KEYNES, J. M. (1921). *Treatise on Probability*. London: Macmillan.

KHACHIYAN, L. G. (1979). A polynomial algorithm in linear programming, *Doklady Akademii Nauk SSSR* **244**:S, 1093-1096, translated in *Soviet Mathematics Doklady* **20**:1 (1979), 191-194.

KOOPMAN, B.O. (1940). The bases of probability, *Bulletin of the American Mathematical Society*, **46**. Reprinted in H. Kyburg and H. Smokler, eds., *Studies in Subjective Probability*, New York: Krieger, 1980, Second edition.

LEAMER, E. (1986). Bid-ask spreads for subjective probabilities, in P. Goel and A. Zellner (eds.), *Bayesian Inference and Decision Techniques: Papers in Honor of Bruno de Finetti*, Amsterdam: North Holland.

LEVY, I. (1980). *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*, Cambridge: M. I. T. Press.

LITTERMAN, R. (1986). Forecasting with Bayesian vector autoregressions - Five Years of Experience, *Journal of Business and Economic Statistics*, **4**, 25-38.

MCNEES, S. (1986). Forecasting accuracy of alternative techniques: a comparison of U. S. macroeconomic forecasts, with discussion, *Journal of Business and Economic Statistics*, **4**, 5-23.

RAHMAN, M. (1987). *Computable Reformulations and Extensions of de Finetti's Fundamental Theorem of Probability*, Ph.D. thesis in mathematics and statistics, State University of New York at Albany.

REICHENBACH, H. (1949). *The Theory of Probability*. Los Angeles: University of California Press.

RESNIK, M. (1980). *Frege and the Philosophy of Mathematics*. Ithaca: Cornell University Press.

SAVAGE, L. J. (1954). *The Foundations of Statistics*. New York: John Wiley.

SCOTT, D. (1964). Measurement structures and linear inequalities. *Journal of Mathematical Psychology*, **1**, 233-247.

SHACKLE , G. L. S. (1949). *Expectation in Economics*. Cambridge: University Press.

_____ (1955). *Uncertainty in Economics*. Cambridge: University Press.

SHAFER, G. (1976). *A Mathematical Theory of Evidence*. Princeton: University Press.

_____ (1982). The Jeffreys-Lindley paradox (with discussion). *Journal of the American Statistical Association*, **77**, 325-351.

SMITH, C. A. B. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society*, Series B., **23**, 1-25.

SUPPES, P. (1974). The measurement of belief. *Journal of the Royal Statistical Society*, Series B., **36**, 160-175 .

_____ (1981). *Logique du Probable*. Paris: Flammarion.

TAYLOR, R. L., DAFFER, P. Z., and PATTERSON, R. F. (1985). *Limit Theorems for Sums of Exchangeable Random Variables*. Totowa, NJ: Rowman and Allanheld.

VAN HEIJENOORT, J., ed., (1967). *From Frege to Godel: A Sourcebook in the History of Mathematical Logic*. Cambridge Mass.: Harvard University Press.

WALSH, G. R. (1985). *An Introduction to Linear Programming*. New York: John Wiley, Second edition.

WHITEHEAD, A. N. and RUSSELL, B. (1910). *Principia Mathematica*. Cambridge: University Press.

WILLEY, F. C. (1960). The three-decision multiple-choice test: a method of increasing the sensitivity of the multiple-choice item. *Psychological Reports*, **7**, 475-477.